
UofS Pulse Binfo Germplasm Documentation

Release 1

Shen Y, Sanderson LA, Tan R. University of Saskatchewan, Pulse

May 12, 2023

Contents:

| | | |
|----------|--|-----------|
| 1 | Installation | 3 |
| 1.1 | Download Package | 3 |
| 1.2 | Enable Package | 3 |
| 1.3 | Set Permissions | 4 |
| 1.4 | Import Data | 5 |
| 1.5 | Upgrade path from Separate Modules | 5 |
| 2 | Data Import | 7 |
| 2.1 | Germplasm Cross Importer | 7 |
| 2.2 | Germplasm Accession Importer | 8 |
| 3 | Germplasm Collection | 13 |
| 3.1 | Functionality | 13 |
| 4 | Recombinant Inbred Line (RIL) Summary | 17 |
| 4.1 | Functionality | 17 |
| 4.2 | Adding RILs to the summary | 19 |
| 4.3 | Adding the summary to RIL pages. | 19 |

Note: This package of modules enhances the germplasm support in Tripal Core with support for both **Breeding Programs** and **Germplasm Genebanks**.

Note: It is recommended to clear caches regularly in this installation processes.

1.1 Download Package

The package is available as one repository for [Pulse Bioinformatics, University of Saskatchewan](#) on GitHub. Recommended method of downloading and installation is using git:

```
cd [your drupal root]/sites/all/modules
git clone https://github.com/UofS-Pulse-Binfo/kp_germplasm.git
```

1.2 Enable Package

The module can be enabled in “Home » Administration » Tripal » Modules” by select the checkbox under “ENABLED” column (as shown in above image) and then click on “Save Configuration” button by the bottom of page.

TRIPAL GERMLASM

| ENABLED | NAME | VERSION | DESCRIPTION | OPERATIONS |
|-------------------------------------|--|---------|---|------------|
| <input checked="" type="checkbox"/> | Germplasm Collection | | Provides fields and views to support germplasm collections in Tripal. Requires: Tripal (enabled), Views (enabled), Chaos tools (enabled), Path (enabled), Search (enabled), PHP filter (enabled), Entity API (enabled), Redirect (enabled), Tripal Chado (enabled), Date (enabled), Date API (enabled), Image (enabled), File (enabled), Field (enabled), Field SQL storage (enabled), Link (enabled), Tripal Chado Views (enabled) | |
| <input checked="" type="checkbox"/> | KP Germplasm | | Generic support for germplasm stored in chado stock. Requires: Tripal (enabled), Views (enabled), Chaos tools (enabled), Path (enabled), Search (enabled), PHP filter (enabled), Entity API (enabled), Redirect (enabled), Tripal Chado (enabled), Date (enabled), Date API (enabled), Image (enabled), File (enabled), Field (enabled), Field SQL storage (enabled), Link (enabled), Tripal Chado Views (enabled) | |
| <input checked="" type="checkbox"/> | RIL (Recombinant Inbred Line) Summary | | Provides a matrix describing the number of RILs per species combination by looking at parentage. Requires: Tripal (enabled), Views (enabled), Chaos tools (enabled), Path (enabled), Search (enabled), PHP filter (enabled), Entity API (enabled), Redirect (enabled), Tripal Chado (enabled), Date (enabled), Date API (enabled), Image (enabled), File (enabled), Field (enabled), Field SQL storage (enabled), Link (enabled), Tripal Chado Views (enabled) | |

Another method that can enable our module is using drush:

```
drush pm-enable kp_germplasm
drush pm-enable rilsummary germplasmcollection
```

Note: In this step, module required ontologies and controlled vocabularies will be inserted into Chado. **Make sure to run any Tripal jobs created by these modules before continuing.**

1.3 Set Permissions

By default, permission of using both importers in this module is not set. It can be configured in “Home » Administration » People » Permissions”.

| PERMISSION | ANONYMOUS USER | AUTHENTICATED USER | BREEDING PROGRAM: ACCESS | BREEDING PROGRAM: MANAGE | ADMINISTRATOR | LENTIL GENOME ACCESS | CONTENT MANAGER | SEQUENCING | PATHOGEN GENOMICS | PHENOTYPES | UNIVERSITY OF SASKATCHEWAN | PHENOTYPES W/ DOWNLOAD | PRIVILEGED COLLABORATOR (NO DOWNLOAD) |
|---|--------------------------|--------------------------|--------------------------|--------------------------|-------------------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|----------------------------|--------------------------|---------------------------------------|
| <input type="button" value="Reset changes"/> <input type="button" value="Save permissions"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Importer: Use the OBO Vocabulary Loader Allow the user to import data using the OBO Vocabulary Loader. Note: you may also need to give the "Upload Files" permission for importers to work. Warning: Give to trusted roles only; this permission has security implications. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Importer: Use the Germplasm Cross Importer Allow the user to import data using the Germplasm Cross Importer. Note: you may also need to give the "Upload Files" permission for importers to work. Warning: Give to trusted roles only; this permission has security implications. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Importer: Use the Germplasm Accession Importer Allow the user to import data using the Germplasm Accession Importer. Note: you may also need to give the "Upload Files" permission for importers to work. Warning: Give to trusted roles only; this permission has security implications. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

1.4 Import Data

After the module is installed and enabled, both Germplasm Cross Importer and Germplasm Accession Importer should be ready to use in “Home » Administration » Tripal » Data Loader”.



Load germplasm into database

Upload file must in TSV (tab-separated values) format. Please confirm file format and column order before upload.

FILE UPLOAD

Germplasm file should be a tab separated file with the following columns:

1. Year: the year the cross was made in.
2. Season: the season the cross was made in (e.g. Spring, Fall, Winter, Summer).
3. Cross Number: a unique identifier for the cross (e.g. 12345).
4. Maternal Parent: the name of the maternal parent of this cross.
5. Paternal Parent: the name of the paternal parent of this cross.
6. Cross Type: the type of cross (e.g. single, double, triple).
7. Seed Type: either the market class or seed coat colour of the seed resulting from the cross.
8. Cotyledon Colour: the cotyledon colour of the seed resulting from the cross.
9. Comment: a free-text comment about the cross.

File Upload

| FILE | SIZE | UPLOAD PROGRESS | ACTION |
|---|------|-----------------|--------|
| <input type="button" value="Choose file"/> No file chosen | | | |

Remember to click the "Upload" button below to send your file to the server. This interface is capable of uploading very large files. If you are disconnected you can return, reload the file and it will resume where it left off. Once the file is uploaded the "Upload Progress" will indicate "Complete". If the file is already present on the server then the status will quickly update to "Complete".



Load germplasm into database

Upload file must in TSV (tab-separated values) format. Please confirm file format and column order before upload.

FILE UPLOAD

Germplasm file should be a tab separated file with the following columns:

1. Germplasm Name: Name of this germplasm accession.
2. External Database
3. Accession Number: A unique identifier for the accession.
4. Germplasm Genus: The genus of the accession.
5. Germplasm Species: The species of the accession.
6. Germplasm Subtaxa: Subtaxon can be used to store any additional taxonomic identifier.
7. Institute Code: The code for the Institute that has bred the material.
8. Institute Name: The name of the Institute that has bred the material.
9. Country of Origin Code
10. Biological Status of Accession: The 3 digit code representing the biological status of the accession.
11. Breeding Method: The unique identifier for the breeding method used to create this germplasm.
12. Pedigree: The cross name and optional selection history.
13. Synonyms: The synonyms of the accession.

File Upload

| FILE | SIZE | UPLOAD PROGRESS | ACTION |
|--|------|-----------------|--------|
| <input type="button" value="Browse..."/> No file selected. | | | |

Remember to click the "Upload" button below to send your file to the server. This interface is capable of uploading very large files. If you are disconnected you can return, reload the file and it will resume where it left off. Once the file is uploaded the "Upload Progress" will indicate "Complete". If the file is already present on the server then the status will quickly update to "Complete".

For more information on the importers, See the Data Import section of these docs.

Note: The importers add data to from your file into Chado. You then need to publish that data by going to Admin > Content > Tripal Content > Publish and selecting either “F1” for crosses or “Germplasm Accessions”.

1.5 Upgrade path from Separate Modules

This package includes modules which used to stand alone (germ_summary, tripal_germplasm_importer, germcollection). To upgrade,

1. Take note of any existing configuration both in Tripal > Extensions and Structure > Tripal Content Types.

2. Disable and uninstall the existing modules. This will not delete any data in chado; however, you will need to re-configure the functionality.
3. Remove the old module directories.
4. Clone this package and re-install the modules
5. Re-apply the configuration you took note of above.

Warning: You may need to re-configure after upgrading to this package so take careful note of your original configuration.

Note: The functionality from separate modules will still be available in this package and any new functionality will be developed here. Additionally, germ_summary has been renamed to rilsummary to reflect it's focus on RILs.

Import for **Germplasm Accessions** and **Breeding Crosses** are currently supported by this module. The accession import supports BrAPI-compliant metadata in a simple table-based format. For more information, see the following full descriptions of each importer.

2.1 Germplasm Cross Importer

2.1.1 Prepare a Germplasm Cross File

Germplasm Cross Importer allows bulk load germplasm crosses into a database. Germplasm cross file needs to follow a specific templet to be able to upload. The following columns must include:

1. **Year:** the year the cross was made in
2. **Season:** the season the cross was made in
 - 2.1 Make sure to have the full name of a season
3. **Cross Number:** a unique identifier for the cross (e.g. 6673S)
 - 3.1 Cross number may already exist in database, double check to make sure the cross number matches exactly the stocked name in database
4. **Maternal Parent:** the name of the maternal parent of this cross
 - 4.1 Name of the maternal parent may already exist in database, double check to make sure the cross number matches exactly the stocked name in database
5. **Paternal Parent:** the name of the paternal parent of this cross
 - 5.1 Name of the paternal parent may already exist in database, double check to make sure the cross number matches exactly the stocked name in database
6. **Cross Type:** the type of cross (e.g. Single Cross, Back Cross)

6.1 Cross type information may be able to find from Cross number. A capitalized letter tends to appear within a cross number, which indicates the Cross type. “S” stands for single cross, “M” stands for multiple cross, “T” stands for triple cross, and “B” stands for back cross. The letter may also be found in low case or missing.

7. **Seed Type:** either the market class or seed coat colour of the progeny
8. **Cotyledon Colour:** the cotyledon colour of the seed resulting from the cross
9. **Comment:** a free-text comment about the cross

Add more columns as needed (e.g. Seed coat, Male Cotyledon Color, Female Cotyledon Color).

2.1.2 Prefix and Organism

Organism must be selected from dropdown menu before upload. Prefix text box is optional to fill in and default the value is ‘GERM’. The unique name for each germplasm will be ‘GERM’ followed by its `stock id` but user can give a unique prefix to replace ‘GERM’.

Prefix

Please give a prefix for this germplasm cross. Unique names for germplasm contained in file will be this prefix followed by stock id. Unique prefix can help you track this germplasm in the future, and default prefix is GERM.
NOTE: unique names for existed germplasm cross will be updated with new prefix too.

Organism *

- Select -

Select the organism for germplasm file you would like to upload

Import File

2.1.3 Bulk load germplasm crosses

As Chado is the data store for Tripal, germplasm crosses will be saved in five `chado tables`: `cv`, `cvterm`, `stock`, `stockprop`, and `stock_relationship` in this module.

- required control vocabularies (CVs) and CV terms will be checked before data loading
- germplasm crosses will be loaded into table `stock`
- properties for each germplasm will be loaded into table `stockprop`
- relationships with parents for each germplasm will be loaded into table `stock_relationship`

2.2 Germplasm Accession Importer

2.2.1 File Upload

Format requirements for upload files can be found easily in UI while using this module.

FILE UPLOAD

Germplasm file should be a tab separated file with the following columns:

1. Germplasm Name: Name of this germplasm accession.
2. External Database: The institution who assigned the following accession.
3. Accession Number: A unique identifier for the accession.
4. Germplasm Genus: The genus of the accession.
5. Germplasm Species: The species of the accession.
6. Germplasm Subtaxa: Subtaxon can be used to store any additional taxonomic identifier.
7. Institute Code: The code for the Institute that has bred the material.
8. Institute Name: The name of the Institute that has bred the material.
9. Country of Origin Code: 3-letter ISO 3166-1 code of the country in which the sample was originally.
10. Biological Status of Accession: The 3 digit code representing the biological status of the accession.
11. Breeding Method: The unique identifier for the breeding method used to create this germplasm.
12. Pedigree: The cross name and optional selection history.
13. Synonyms: The synonyms of the accession.

Note: For column 2 (External Database), the name in file must exist in your database in chado table:db already.

For column 12 (Pedigree), pedigree information is recommended to save in format of maternal-parent-name/paternal-parent-name.

For column 13 (Synonyms), multiple synonyms are allowed but must be separated by semi-colons (e.g. syn1; syn2; syn3).

2.2.2 Select Genus

Genus of the accessions in file must be selected from dropdown menu before upload. All accessions in one file must belong to same genus and match this selection.

File Upload

| FILE | SIZE | UPLOAD PROGRESS | ACTION |
|-----------------------------|------|-----------------|--------|
| Browse... No file selected. | | | |

Remember to click the "Upload" button below to send your file to the server. This interface is capable of uploading very large files. If you are disconnected you can return, reload the file and it will resume where it left off. Once the file is uploaded the "Upload Progress" will indicate "Complete". If the file is already present on the server then the status will quickly update to "Complete".

Upload File

Server path

If the file is local to the Tripal server please provide the full path here.

Remote path

If the file is available via a remote URL please provide the full URL here. The file will be downloaded when the importer job is executed.

Genus *

- Select -

Select the genus for germplasm accession file you would like to upload

Import File

2.2.3 Bulk load germplasm accessions

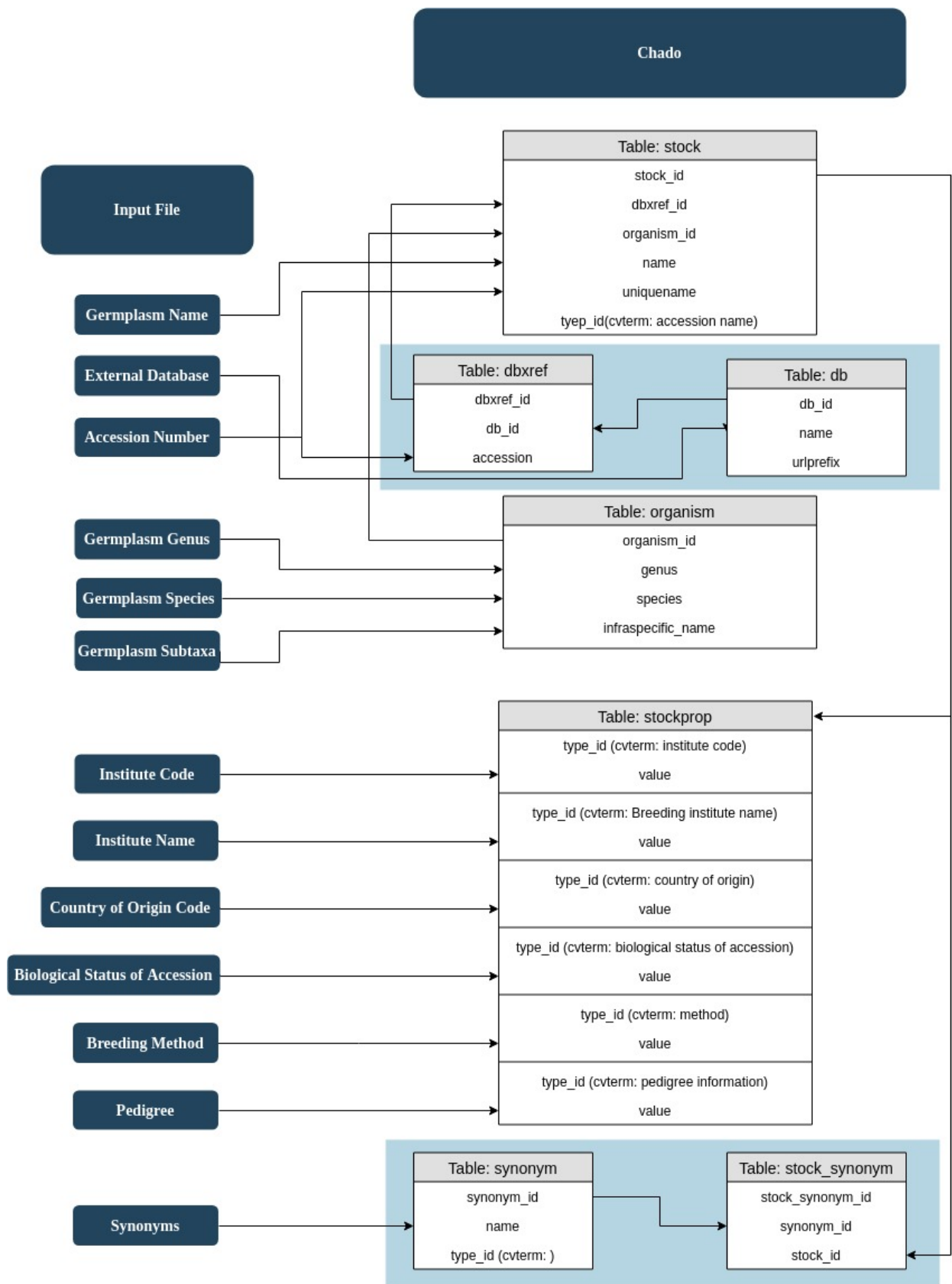
As Chado is the data store for Tripal, germplasm accessions will be saved in several chado tables: cv, cvterm, stock, stockprop, db, dbxref, synonym and stock_synonym for this importer.

The general idea of how accession information will be saved in database:

- required control vocabularies (CVs) and CV terms will be checked before data loading

- organism of one accession is determined by germplasm genus, species and subtaxa(optional)
- germplasm accession will be loaded into table stock, dexref, and db
- properties will be loaded into table stockprop
- synonyms will be loaded into table synonym and stock_synonym

A diagram:



Note: This module uses a specific set of controlled vocabulary terms to identify metadata. Where possible we have used community standard ontologies but in some cases the terms needed were not available. All terms used are compatible with standardized Tripal Content Types.

Germplasm Collection

Provides functionality for supporting germplasm collection (e.g. diversity panels and recombinant inbred lines) including the following:

- Germplasm Collection content type is created automatically.
- **Fields:**
 - Germplasm List (co_010__germplasm): Listing of all germplasm in a collection.
 - Project-Related Germplasm Collection (local__project_germcollection): to link germplasm collections with projects.

3.1 Functionality

3.1.1 Germplasm Collection Pages

This module creates a germplasm collection content type which you can configure at Admin > Structure > Tripal Content Types > Germplasm Collection. By default, there will be a name, identifies, type of collection and germplasm list. The germplasm list will show all chado.stock records linked to a given chado.stockcollection record via the chado.stockcollection_stock table in an ajax-paged list.

Sunny lozenge goomer (SLGcollection)

View

Edit

Reload

[Summary](#)
[Germplasm](#)

| Summary | |
|-----------------------------|--------------------------------|
| Resource Type | Germplasm Collection |
| Name | Sunny lozenge goomer |
| Identifier | SLGcollection |
| Stockcollection Type | dicistronic_primary_transcript |

| Germplasm | | | |
|----------------------------|----------------------------|---------------------|-----------------|
| Name | Accession | Type | Species |
| SLG006F1 | SLG006F1 | Germplasm Accession | Citrus sinensis |
| SLG006N001 | SLG006N001 | Germplasm Accession | Citrus sinensis |
| SLG006N003 | SLG006N003 | Germplasm Accession | Citrus sinensis |
| SLG006N004 | SLG006N004 | Germplasm Accession | Citrus sinensis |
| SLG006N007 | SLG006N007 | Germplasm Accession | Citrus sinensis |
| SLG006N008 | SLG006N008 | Germplasm Accession | Citrus sinensis |
| SLG006N010 | SLG006N010 | Germplasm Accession | Citrus sinensis |
| SLG006N013 | SLG006N013 | Germplasm Accession | Citrus sinensis |
| SLG006N014 | SLG006N014 | Germplasm Accession | Citrus sinensis |
| SLG006N015 | SLG006N015 | Germplasm Accession | Citrus sinensis |

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 next › last »

Warning: There is currently no way to link a Germplasm collection with germplasm (chado.stock) through the user interface. Instead, you will need to add records to chado.stockcollection_stock manually to create the link.

3.1.2 Link Germplasm to Projects

There is also a field, Project-Related Germplasm Collection (local__project_germcollection), which links germplasm collections to projects. This can be used to indicate which germplasm are used for a given experiment. The following example shows two germplasm collections linked to a single project.

Test project 1 with germplasm collection

View

Edit

Reload

[Summary](#)
[Publication](#)
[Relationship](#)

| Summary ✕ | |
|-----------------------------|--|
| Resource Type | Project |
| Name | Test project 1 with germplasm collection |
| Short Description | |
| Germplasm Collection | Sunny lozenge goomer Sunny Ground Loomers |

Note: You can add germplasm collections to projects by editing a project and entering the germplasm collection name.

Recombinant Inbred Line (RIL) Summary

Provides functionality for summarizing Recombinant Inbred Lines (RILs) including the following:

1. Tabular matrix which summarizes how many RILs are available for each species combination. This is particularly helpful if you have a cultivated and associated wild species for a single genus.
2. Listing of all RILs for a specific species combination including information about the number of F2 families for each F-generation.
3. ChadoField for RIL pages which summarizes information about the number of F2 families for each F-generation.

4.1 Functionality

The RIL summary matrix can be found at `[mytripalsite.com]/germplasm/summary/[genus]`. This is what it looks like for a fake *Tripalus* example.

Tripalus Germplasm Summary

The following table summarizes the number of germplasm with parents of the specified species. Specifically, the maternal parent species is listed as the column header and the paternal parent species is the row header. Thus intraspecific crosses are shown along the rest of the table.

| | | Maternal Parent | | |
|-----------------|-------------------|-------------------|-------------------|-----------------|
| | | <i>databasica</i> | <i>drupalicus</i> | <i>biodatum</i> |
| Paternal Parent | <i>databasica</i> | 2 | 0 | 0 |
| | <i>drupalicus</i> | 0 | 0 | 0 |
| | <i>biodatum</i> | 0 | 0 | 0 |

When you click on any of the cells in the RIL summary matrix you are taken to the following listing:

Tripalus Germplasm Summary

datatabaseica
Maternal Parent

datatabaseica
Paternal Parent

2 Stock Records

[Back to Summary Matrix](#)

| Name | Accession | Maternal Parent | Paternal Parent | Generations | | | | | | | | <input checked="" type="checkbox"/> | |
|-------|-----------|---------------------|-------------------|-------------|----|----|----|----|----|----|----|-------------------------------------|--|
| | | | | F1 | F2 | F3 | F4 | F5 | F6 | F7 | F8 | | |
| TR-01 | TR-01 | CDC FRED (CDC FRED) | AABC-12 (AABC-12) | | | | | | | | | | |
| TR-02 | TR-02 | CDC FRED (CDC FRED) | AABC-12 (AABC-12) | | | | | | | | | | |

The details for a given RIL can be summarized on the RIL Tripal Content Page using the field provided by this module.

TR-02

View

Edit

Reload

Summary
 Annotations
 Cross
 Reference
 Publication
 Relationship

Summary ✕

| | |
|----------------------|--|
| Resource Type | Recombinant Inbred Line |
| Accession | |
| Organism | Tripalus biodatum |
| Name | TR-02 |
| Identifier | TR-02 |
| Description | |
| Is Obsolete | No |

F1
F2
F3
F4
F5
F6
F7
F8

RIL Complete

Values indicate number of sublines in a generation.

TR-02 (TR-02) is 414 inbred line
 Maternal and Paternal Parents: CDC FRED/AABC-12

4.2 Adding RILs to the summary

1. Create a Recombinant Inbred Line with the name of your RIL population (e.g. TR-01).
2. Create a germplasm line (type doesn't matter; suggested Generated Germplasm (Breeding Line)) with the name of the original cross giving rise to the RIL population (e.g. 1234S) and add a relationship: TR-01 is_selection_of 1234S.
3. Create parents for the Breeding cross (type does not matter) and related them using the is_maternal_parent and is_paternal_parent relationship types (e.g. CDC FRED is_maternal_parent_of 1234S and AABC is_paternal_parent_of 1234S).
4. Each subline for a RIL (i.e. TR-01-123) should be of type *stock_type:F2*.

4.3 Adding the summary to RIL pages.

1. Go to Admin > Structure > Tripal Content Types > Recombinant Inbred Lines > Manage Fields.
2. Add a new field where the type is Germplasm RIL Summary.

3. Make sure it is not disabled on the Manage Display tab.