
Tripal Genetic Documentation

LA Sanderson (University of Saskatchewan, Pulse Bioinformatics)

Apr 21, 2020

Contents:

1	Installation	3
1.1	Dependencies	3
1.2	Installation	3
2	MSTmap Importer	5
2.1	Format	5
2.2	Validation	6
2.3	Data Storage	7
3	QTL Importer	9
3.1	Format	9
3.2	Validation	10
3.3	Data Storage	10

This module provides additional fields and data import which provide support for quantitative trait loci. Additionally, it integrates with Tripal Map through shared vocabulary terms and data storage models.

1.1 Dependencies

- Tripal 3.x

1.2 Installation

The preferred method of installation is using Drush:

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

The above command downloads the module into the expected directory (e.g. `/var/www/html/sites/all/modules/tripal_qtl`). Next we need to install the module:

```
drush pm-enable tripal_qtl
```

That's it! No configuration needed!

2.1 Format

This importer will load genetic maps which follow the same format output by the [MSTmap software](#). This format consists of a TSV file listing the positions of markers in a genetic map. Each linkage group is labelled with it's name followed by a 2-column format: marker name, marker position (centimorgans). For example,

```
group lg0
m4 0.000
m2 0.000
m3 0.000
m1 0.000
m8 4.009
m12 4.009
m6 4.009
m7 4.009
m9 5.009
```

The importer also stores metadata about the map through a well described form:

FILE UPLOAD

MSTmap format consists of a TSV file listing the positions of markers in a genetic map. Each linkage group is labelled with it's name followed by a 2-column format: marker name, marker position (centimorgans). For example,

```
group 1g0
m4 0.000
m2 0.000
m3 0.000
m1 0.000
m8 4.009
m12 4.009
m6 4.009
m7 4.009
m9 5.009
```

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".

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.

Map Name *

A unique canonical name for the linkage map as defined by the curator.

Published Map Name

The name of the map as it was published.

Map Species *

The species the map should be categorized under based on the parental species. Choose interspecific if the parental species are different.

Map Type *

The type of map.

Population Type

The type of population used to generate the map (e.g. F2, BC1, RIL, ABC).

Population Size

The number of individuals in the population used to generate this map.

Mapping Software *

The name of the mapping software used to generate the map (e.g. MSTmap, MapMaker EXP, MapMaker Macintosh, Mapdisto, CarthaGene, MergeMap, QTL IciMapping, JoinMap)

Mapping Software Version *

The version of the above mapping software which was used.

Methodology

Description of how the linkage map was created. Specifically, include software parameters, whether the markers were binned and describe any manual manipulation of the map.

Map Description

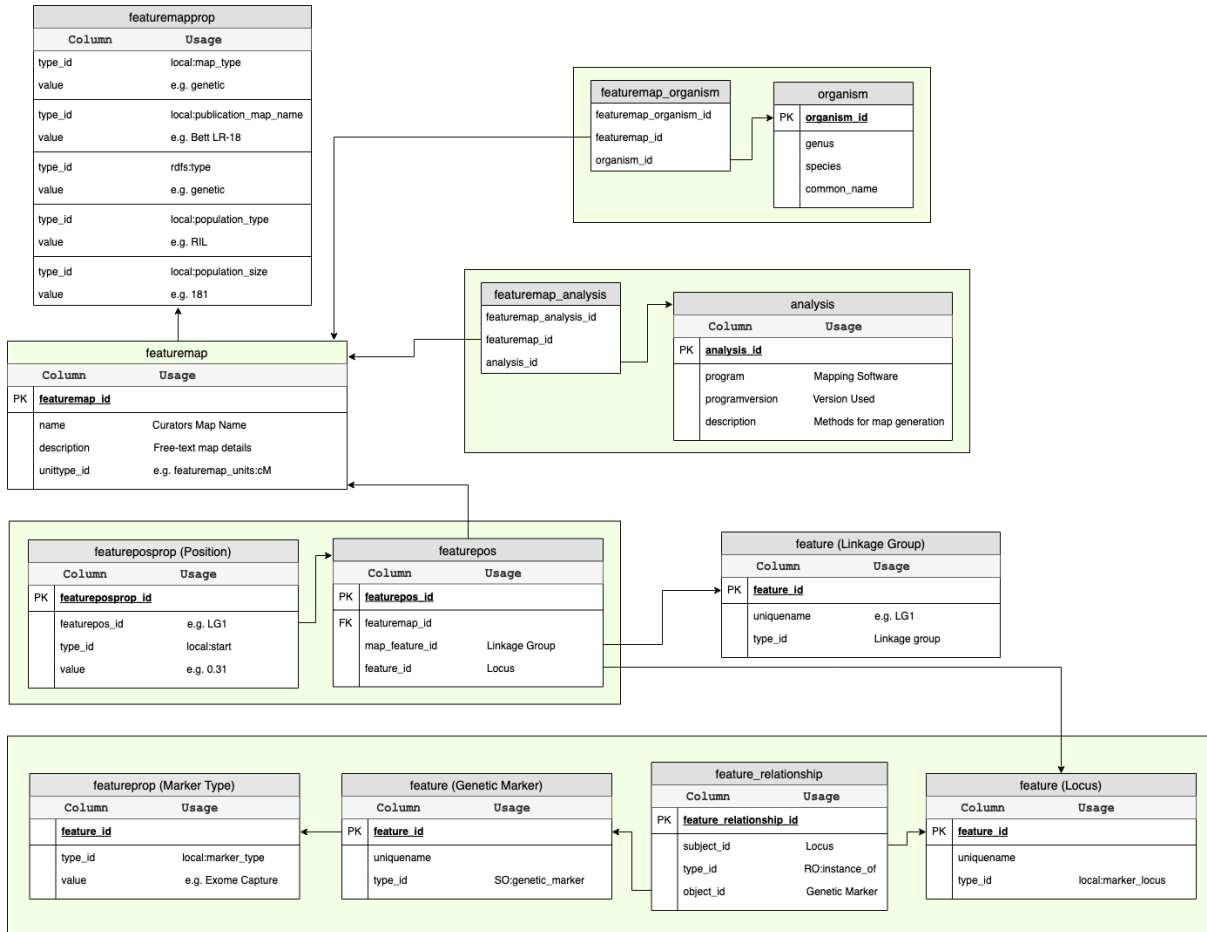
Description of the linkage map. Suggested topics include the published source; alternative names for the map; types of molecular markers; parents, size, and type of mapping population; linkage group number, length, marker count, and sub-genome assignment; the map's total length and marker count; average inter-marker gap distance or marker density. Indicate if there are PCR primers that amplify more than a single locus on the map, if there is a framework map available in addition to the full genetic map, and compare and contrast with other genetic maps. Describe whether linkage groups are homeologous and to what extent common markers and conserved marker orders are present. Record if the map was produced directly using a bi-parental mapping population or whether it is a consensus or composite map; if it is the latter then list and describe briefly the constituent mapping populations used in its construction. Describe how many markers were screened against the parents, what number were polymorphic and used for population genotyping, and what proportion appeared on the final map.

2.2 Validation

There is currently no validation of the file on import. However, the metadata in the form is minimally validated (e.g. the organism must already exist in order to be chosen).

2.3 Data Storage

Metadata about the map including it's name, description, population type and size are stored in the featuremap and featuremapprop tables of chado; whereas, metadata describing the methodology is stored in the analysis table. Each position in the map is a combination of the linkage group, locus and genetic marker (features) with the position itself stored in the featurepos table. This is more clearly explained via the ER diagram below.



This importer will load quantitative trait loci (QTL) and associate them with a pre-existing map.

Warning: You must pre-load the map using the MST Map importer before you can load QTL associated with it.

3.1 Format

We have developed our own format for this importer due to inconsistency in the community. The file must be tab-delimited, have linux line endings and a header. The description of the columns is as follows:

1. QTL Published Name: The name of the QTL as it was published.
2. Trait Full Name: The full name of the trait the QTL is contributing to (e.g. Days to Flower).
3. Trait Abbreviation: A short 2-5 letter abbreviation of the trait (e.g. DTF).
4. Environments (SiteYears): The environments a QTL is significant in (e.g. Preston 2009).
5. Peak Marker: The genetic marker closest to the peak LOD of the QTL.
6. Linkage Group: The linkage group the QTL is located on.
7. Peak Position: The cM position of the peak LOD for the QTL.
8. Peak LOD: The highest LOD score on the QTL graph.
9. Phenotypic r^2 : Phenotypic variance explained by the QTL (expressed as a fraction).
10. Additive Effect: The additive effect of each allele expressed as a fraction. The parent contributing the allele is supplied in the next column. This value is usually supplied as a negative or positive value dependant upon the contributing parent.
11. Contributor Parent: The parent which contributes to the effect of the QTL based on the value of the additive effect.

12. Confidence Interval Left (1 LOD drop): The LEFT cM position on the QTL graph where the LOD is ONE less than the peak LOD.
13. Confidence Interval Right (1 LOD drop): The RIGHT cM position on the QTL graph where the LOD is ONE less than the peak LOD.
14. Confidence Interval Left (2 LOD drop): The LEFT cM position on the QTL graph where the LOD is TWO less than the peak LOD.
15. Confidence Interval Right (2 LOD drop): The RIGHT cM position on the QTL graph where the LOD is TWO less than the peak LOD.

3.2 Validation

- The genetic map must already exist in order for you to choose it in the form.
- The CV for the traits must already exist.
- The Trait name in your file must match a pre-existing term in the chosen trait controlled vocabulary.
- Empty lines are skipped.
- Linkage groups must already exist and names must match exactly what is in the file.
- The confidence interval values are optional.

3.3 Data Storage

QTL data is stored as expected by Tripal Map. Specifically, it is assumed the map matches the format described for the MST Map importer and QTL features are located on that map using the featurepos table. The position stored in featurepos is the position of the QTL peak. This position is also stored in the featureposprop table along with the confidence interval values. See the following ER diagram for the full description of how the data is stored.

