Tripal Genetic Documentation

LA Sanderson (University of Saskatchewan, Pulse Bioinformatics

Apr 21, 2020

Contents:

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.

CHAPTER 1

Installation

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!

CHAPTER 2

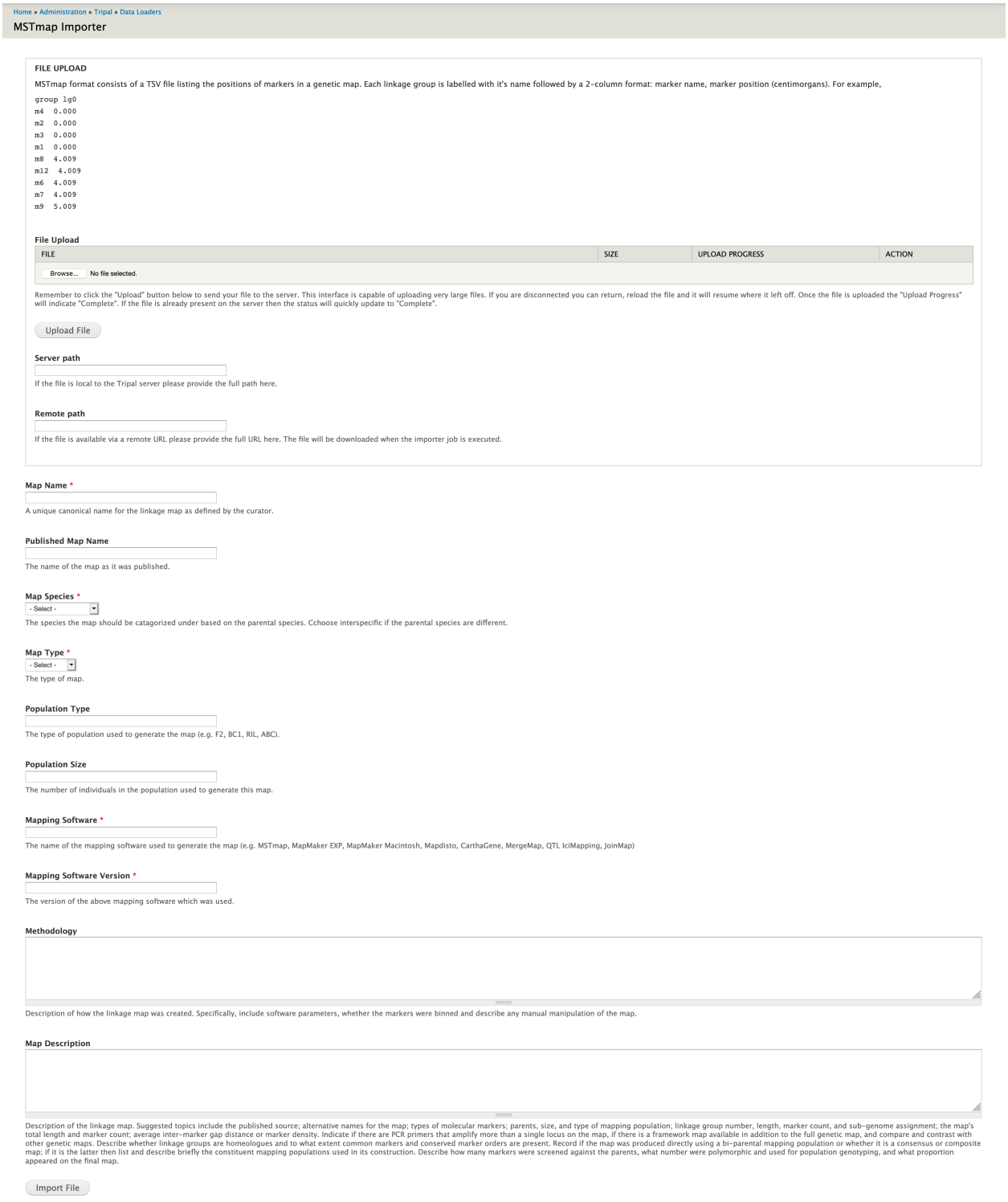
MSTmap Importer

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:

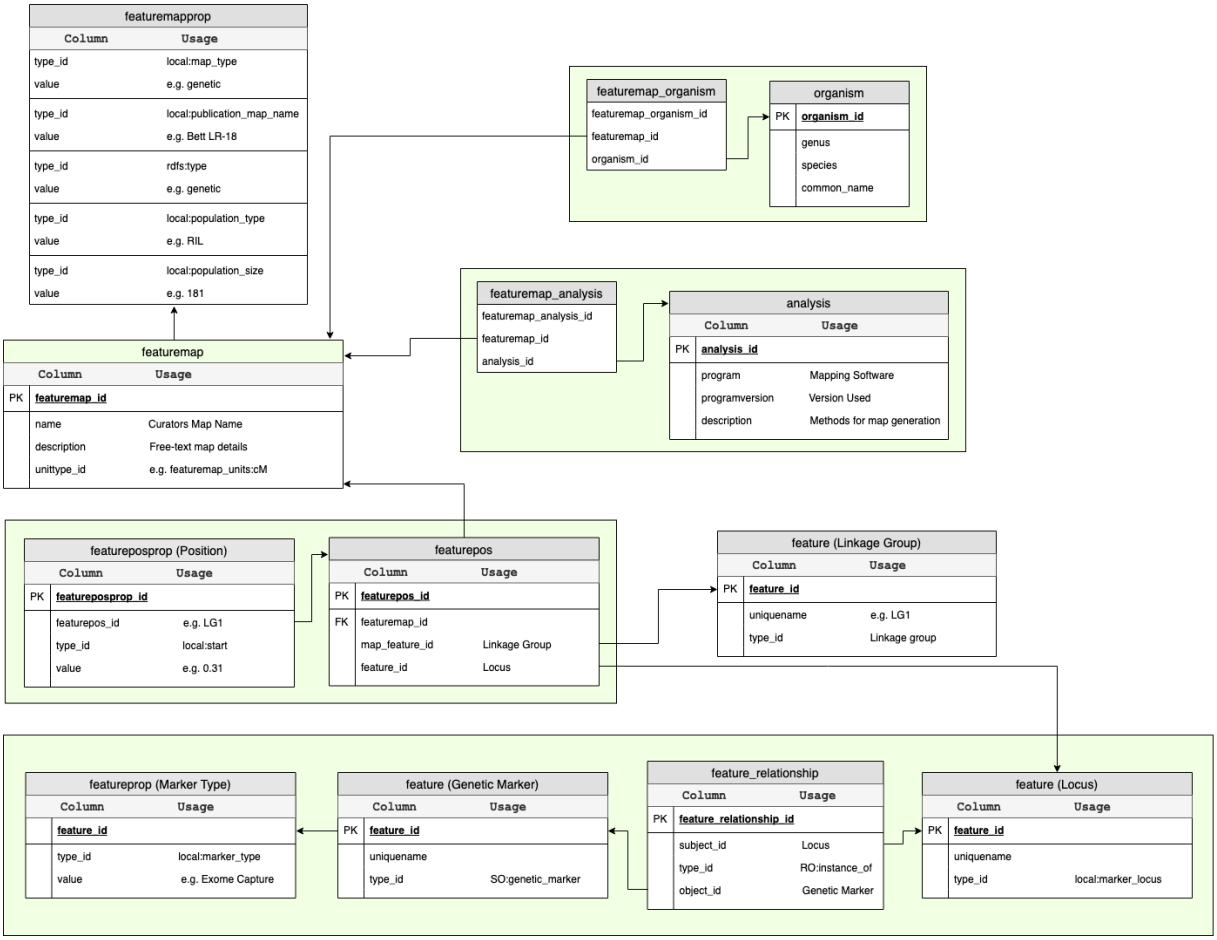


2.2 Validation

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.



CHAPTER 3

QTL Importer

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 r2: 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.

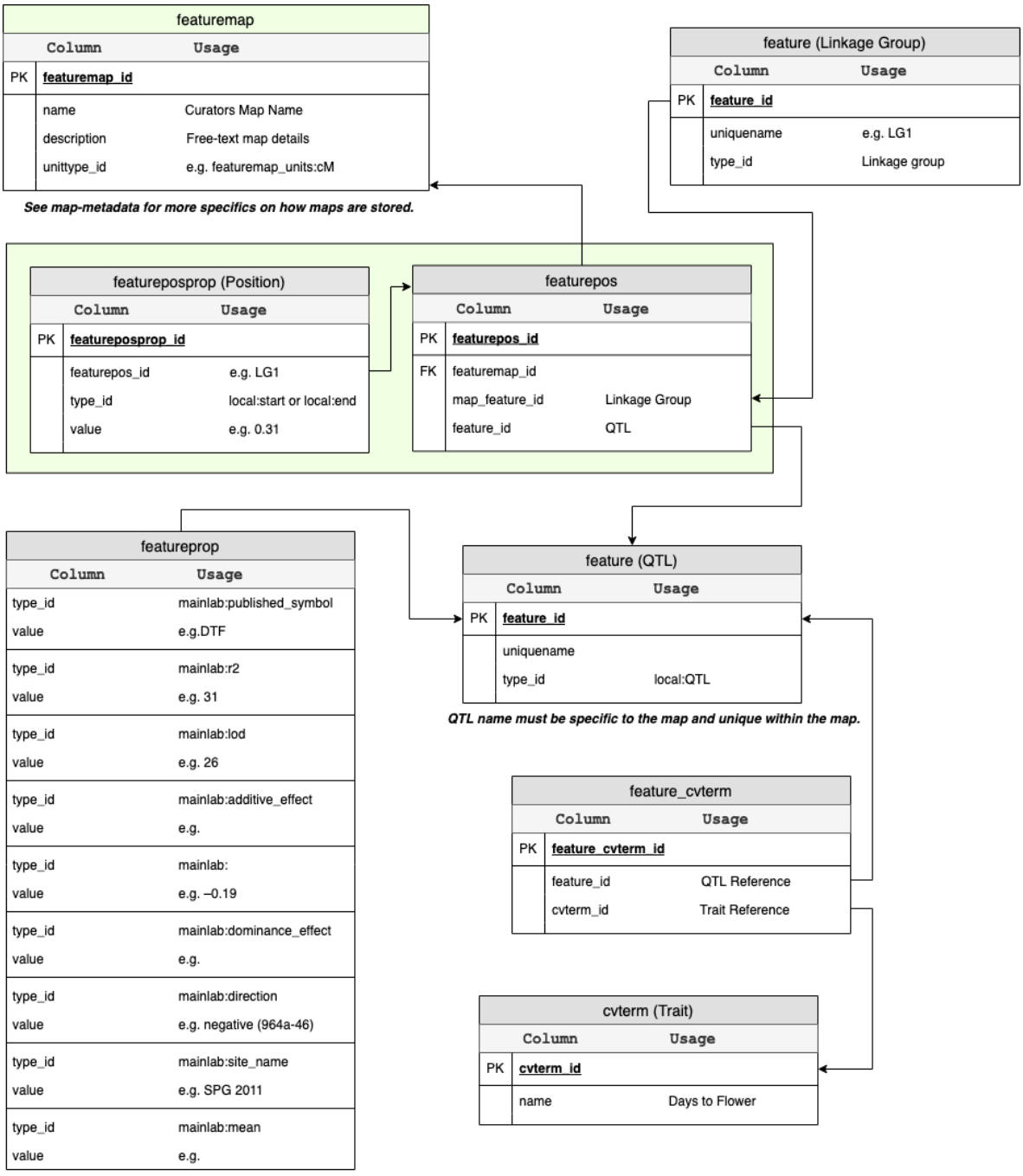
12. Confidence Interval Left (1 LOD drop):The LEFT cM position on the QTL graph where the LOD is ONE less then the peak LOD.
13. Confidence Interval Right (1 LOD drop): The RIGHT cM position on the QTL graph where the LOD is ONE less then the peak LOD.
14. Confidence Interval Left (2 LOD drop): The LEFT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


featuremap
Column Usage
PK featuremap_id
name Curators Map Name
description Free-text map details
unittype_id e.g. featuremap_units:cM
See map-metadata for more specifics on how maps are stored.
feature (Linkage Group)
Column Usage
PK feature_id
uniquename e.g. LG1
type_id Linkage group
featureposprop (Position)
Column Usage
PK featureposprop_id
featurepos_id e.g. LG1
type_id local:start or local:end
value e.g. 0.31
featurepos
Column Usage
PK featurepos_id
FK featuremap_id
map_feature_id Linkage Group
feature_id QTL
featureprop
Column Usage
type_id mainlab:published_symbol
value e.g.DTF
type_id mainlab:r2
value e.g. 31
type_id mainlab:lod
value e.g. 26
type_id mainlab:additive_effect
value e.g.
type_id mainlab:
value e.g. –0.19
type_id mainlab:dominance_effect
value e.g.
type_id mainlab:direction
value e.g. negative (964a-46)
type_id mainlab:site_name
value e.g. SPG 2011
type_id mainlab:mean
value e.g.
feature (QTL)
Column Usage
PK feature_id
uniquename
type_id local:QTL
QTL name must be specific to the map and unique within the map.
feature_cvterm
Column Usage
PK feature_cvterm_id
feature_id QTL Reference
cvterm_id Trait Reference
cvterm (Trait)
Column Usage
PK cvterm_id
name Days to Flower