

---

# **TPPSc - Internal TPPS Pipeline Documentation**

**Peter Richter**

**Feb 10, 2023**



---

## Contents:

---

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Tutorial</b>	<b>3</b>
2.1	Creating a New TPPSc Submission . . . . .	3
2.2	Author and Species Information (Page 1) . . . . .	3
2.3	Experimental Conditions (Page 2) . . . . .	5
2.4	Tree Accession (Page 3) . . . . .	6
2.5	Genotype, Phenotype, and Environmental Data (Page 4) . . . . .	6
<b>3</b>	<b>Features</b>	<b>13</b>
<b>4</b>	<b>Installation</b>	<b>15</b>



# CHAPTER 1

---

## Introduction

---

TPPSc is a [Drupal](#) module built to extend the functionality of the [Tripal](#) toolset. The purpose of this Tripal module is to permit internal curation of manuscripts available on public sites with supplemental data, such as Dryad. This module will allow members of the database team to load existing flat files and select the data that is relevant.



TPPSc is a multi-page form that serves as the internal pipeline to submit scientific data to our database. This tutorial is meant to show users how to successfully use TPPSc to upload data to your database, along with how to specifically use certain features, and what to do depending on what type of study it is.

## 2.1 Creating a New TPPSc Submission

Here you have the opportunity to create a new TPPSc Submission and the chance to continue working on any unfinished submissions. Simply use the dropdown to select your option and proceed to the next page.

Welcome to TPPSC!

If you would like to submit your data, you can click the button 'Continue to TPPSC' below!

Would you like to load an old TPPSC submission, or create a new one?

Create new TPPSC Submission ▼

CONTINUE TO TPPSC

## 2.2 Author and Species Information (Page 1)

This page requires you to enter a dryad DOI accession number of the study you plan on uploading. The dryad DOI should pull the species, authors, abstract, title, and journal of the paper. This allows the user to enter only one textfield to pull all relevant information.

⚠ Author and Species Information
✖ Experimental Conditions
✖ Tree Accession
✖ Submit Data

DOI: \*

Example: 123.456/dryad.789

**▼ ORGANISM INFORMATION:**

Please provide the name(s) of the species included in this publication.

Species 1: \*

\*: Required Field

Note, you do not enter the DOI, but the dryad DOI, and not the entire link, but only past the colon. If the full dryad DOI link was <https://datadryad.org/resource/doi:10.5061/dryad.5jn04>, then only enter 10.5061/dryad.5jn04.

⚠ Author and Species Information
✖ Experimental Conditions
✖ Tree Accession
✖ Submit Data

DOI: \*

Example: 123.456/dryad.789

The publication has been successfully loaded from Dryad:

Title: Data from: Worldwide core collections of tea (Camellia sinensis) based on SSR markers

Primary Author: Taniguchi, Fumiya

**▼ ORGANISM INFORMATION:**

Please provide the name(s) of the species included in this publication.

Species 1: \*

\*: Required Field

To check if it successfully pulled all relevant information, theres a green right hand bar that updates as you move throughout the form. Below is an example of how it should look once you move to the next page.



✔ Step 1 - Complete  
Author and Species Information

✔ Publication Title: Data from:  
Worldwide core collections of tea  
(Camellia sinensis) based on  
SSR markers

✔ Primary Author: Fumiya  
Taniguchi

✔ Publication Journal: Tree  
Genetics & Genomes

✔ Secondary Author 1: Keiko  
Kimura

✔ Secondary Author 2: Tetsuji  
Saba

✔ Secondary Author 3: Akiko  
Ogino

✔ Secondary Author 4: Satoshi  
Yamaguchi

✔ Secondary Author 5: Junichi  
Tanaka

✔ Species 1: Camellia sinensis

## 2.3 Experimental Conditions (Page 2)

This page consists of two dropdowns that directly affects pages 3 and 4. The Data Type dropdown is where you chose what type of scientific data you are uploading, whether it be Genotype, Phenotype, Environment, or any combination of the 3. The Study Type dropdown is where you chose what kind of study you are uploading. The options for this field are Natural Population, Growth Chamber, Greenhouse, Experimental/Common Garden, and Plantation. The Data Type selection directly affects page 4 and the Study Type selection directly affects page 3.

✔ Author and Species Information   ⚠ Experimental Conditions   ✖ Tree Accession   ✖ Submit Data

### STUDY DESIGN

Data Type: \*

Study Type: \*

\* : Required Field


BACK

SAVE

NEXT

## 2.4 Tree Accession (Page 3)

This page is where we upload the tree location information. We do this by uploading an excel file that has all of the location information. Once the file is uploaded, you then select which columns represent which required data, where the required data is TreeID, Latitude, and Longitude (and if there are multiple species, then a column that represents the genus and species).

 treeaccessionTUTORIAL.xlsx
 REMOVE

File Upload empty field:

By default, TPPS will treat cells with the value "NA" as empty. If you used a different empty value indicator, please provide it here.

▼ DEFINE DATA

*Please define which columns hold the required data: Tree Identifier and Location. If your trees are located based on a population group, you can provide the population group column and a mapping of population group to location below.*

TreeID	Tree Identifier ▼	Longitude	Longitude ▼	Latitude	Latitude ▼
PAAB0001		10.940116		46.330105	
PAAB0002		10.940393		46.32967	
PAAB0003		10.940557		46.329555	

☐ My file has no header row

Coordinate Projection

Below is an example tree location file:

	A	B	C	
1	TreeID	Longitude	Latitude	
2	PAAB0001	10.940116	46.330105	
3	PAAB0002	10.940393	46.32967	
4	PAAB0003	10.940557	46.329555	
5	PAAB0004	10.940811	46.329778	
6	PAAB0005	10.941132	46.329453	
7	PAAB0006	10.94097	46.329386	
8	PAAB0007	10.940981	46.329205	
9	PAAB0008	10.940644	46.32936	
10	PAAB0009	10.940627	46.329424	
11	PAAB0010	10.940267	46.329551	
12	PAAB0011	10.940185	46.329587	
13	PAAB0012	10.939674	46.329699	
14	PAAB0013	10.939392	46.329518	
15	PAAB0014	10.938884	46.329607	
16	PAAB0015	10.938713	46.329845	
17				

## 2.5 Genotype, Phenotype, and Environmental Data (Page 4)

Page 4 will change depending on what you selected for the Data Type dropdown. If you selected an option with Genotype, then the Genotype fields will be available. The same goes for phenotype and Environment.

## 2.5.1 Genotype

The main two types of genotypic data that TPPSc handles is SNP data, and SSR data. For both situations, there is a dropdown field that asks if a reference genome is used.

**SNP:** When dealing with a SNP study, first you select what kind of SNP's you have by selecting which kind of experimental design it is. The possible dropdowns for Experimental Design are GBS, Targeted Capture, Whole Genome Resequencing, RNA-Seq, and Genotyping Array.

- IF GBS is selected, another dropdown for GBS Type appears with options RADSeq, ddRAD-Seq, NextRAD, and RAPTURE
- IF Targeted Capture is selected, another dropdown for Targeted Capture Type appears with options of Exome Capture or Other

If there is SNP data to be uploaded along with the options above, then checking the SNPs Genotype Assay box will provide you with another excel sheet uploader identical to the Tree Accession, where you now select the columns for TreeID and SNP data. If there was an Assay Design file, there is now an option to upload that as well.

▼ GENOTYPE INFORMATION:

Marker Type (select all that apply): \*

☒ SNPs
 ☐ SSRs/cpSSRs
 ☐ Other

▼ SNPS INFORMATION:

Define Experimental Design: \* - Select - ▼

Reference Assembly used: \* - Select - ▼

Genotype File Types (select all that apply): \*
 

☐ SNPs Genotype Assay
 ☐ VCF

Below is an example SNP data file:

	A	B	C	D
1	SampleID	UMN_853_01_Paab_38	UMN_686_01_Paab_73	UMN_5384_02_Paab_83
2	PAAB0001	GG	AA	GG
3	PAAB0002	GG	GG	GG
4	PAAB0003	N/A	GG	GG
5	PAAB0004	GG	AA	GG
6	PAAB0005	N/A	GG	GG
7	PAAB0006	AA	GG	AG
8	PAAB0007	GG	GG	AG

**SSR:** When dealing with an SSR study, first you describe what type of SSR/cpSSR you have by entering it into the textfield provided. Another field labeled Ploidy then appears which is very important when uploading the the SSR/cpSSR spreadsheet. Since SSR's would be different depending on which allele you are looking at, we need to add in a field to determine the ploidy to know which columns denote what. If an organism is diploid, then there would be two columns for each SSR. If an organism is triploid, then there would be three columns for each SSR.

▼ GENOTYPE INFORMATION:

Marker Type (select all that apply): \*

- ☐ SNPs  
☒ SSRs/cpSSRs  
☐ Other

Define SSRs/cpSSRs Type: \*

Reference Assembly used: \*

- Select - ▼

Ploidy - Select - ▼

Genotype File Types (select all that apply): \*

- ☐ SSRs/cpSSRs Genotype Spreadsheet  
☐ VCF

Below is an example SSR data file for 3 SSR markers where the ploidy would be diploid:

	A	B	C	D	E	F	G
1	Sample ID	TTC1.7		POC55		POC81	
2	AE44	126	126	223	236	101	103
3	AE45	126	126	222	222	103	103
4	AE46	126	129	222	222	103	103
5	AE47	129	129	240	240	103	103
6	AE48	126	126	225	233	103	103
7	AE49	126	126	222	244	97	131
8	AE50	126	126	222	231	101	101
9	AE51	126	129	222	223	105	105

## 2.5.2 Phenotype

When dealing with recorded phenotypes, we create phenotypes in tppsc, and then upload a phenotype file which uses corresponding phenotypes. Each phenotype has 4 different values: Name (human readable name, must match column name exactly), Attribute (thing the phenotype is describing: age, height, amount, etc), Description, and Units. Note, if a user has 20 + phenotypes, they can enter them all by creating an excel file with those 4 columns (name, attribute, description, and units).

ADD PHENOTYPE

REMOVE PHENOTYPE

**Phenotype 1:**

Phenotype 1 Name: \*  ○

Phenotype "name" is the human-readable name of the phenotype, where "attribute" is the thing that the phenotype is describing. Phenotype "name" should match the data in the "Phenotype Name/Identifier" column that you select in your Phenotype file below.

---

Phenotype 1 Attribute: \*  ○

Some examples of attributes include: "amount", "width", "mass density", "area", "height", "age", "broken", "time", "color", "composition", etc.

---

Phenotype 1 Description: \*

Please provide a short description of Phenotype 1

---

Phenotype 1 Units: \*  ○

Some examples of units include: "m", "meters", "in", "inches", "Degrees Celsius", "°C", etc.

---

After creating each phenotype, you now need to upload the file. There are two types of file formats that the phenotype file can be in. File type 1 has all the phenotypes in their own columns, while file type 2 has 3 total columns: TreeID, phenotype, and value. Either are accepted, and once uploaded the user will need to chose which columns represent what, just as done previously.

Phenotype file format: \*

	A	B	C
1	tree id	age	rainfall
2	ttps-test-100	59	15
3	ttps-test-101	91	16
4	ttps-test-102	94	10
5	ttps-test-103	60	15
6	ttps-test-104	86	16
7	ttps-test-105	94	19
8	ttps-test-106	80	20
9	ttps-test-107	71	14
10	ttps-test-108	73	18
11	ttps-test-109	61	17
12	ttps-test-110	80	19

☒ Type 1

	A	B	C
1	tree id	phenotype	value
2	ttps-test-100	rainfall	14
3	ttps-test-101	rainfall	18
4	ttps-test-102	rainfall	12
5	ttps-test-103	rainfall	17
6	ttps-test-104	rainfall	18
7	ttps-test-105	rainfall	12
8	ttps-test-106	rainfall	14
9	ttps-test-107	rainfall	18
10	ttps-test-108	rainfall	10
11	ttps-test-109	rainfall	17
12	ttps-test-110	age	51

☐ Type 2

Please select a file format type from the listed options. Below please see examples of each format type.

Phenotype file: Please upload a file containing columns for Tree Identifier, Phenotype Data: \*

No file chosen

File Upload empty field:

By default, TPPS will treat cells with the value "NA" as empty. If you used a different empty value indicator, please provide it here.

☐ My file has no header row

## 2.5.3 Environment

When dealing with environmental data, there are two types of data. You can use Environmental Layers indexed by CartograTree, and/or layers that you collected yourself. Indexed CartograTree layers vary by site, but common ones include Worldclim v.2, PET and Aridity, and Major Soil Types.

### ▼ ENVIRONMENTAL INFORMATION:

☐ I used environmental layers in my study that are indexed by CartograTree.

If the layer you used is not in the list below, then the administrator for this site might not have enabled the layer group you used. Please contact them for more information.

☐ I have environmental data that I collected myself.

When adding layers that you collected yourself, one would follow the exact same process as adding in a phenotype (Name, description, units, and value).

**CARTOGRATREE ENVIRONMENTAL LAYERS: \***☐ Major Soil Types (Conservation Biology Institute)☐ PET and Aridity (CGIARCSI)☐ Canopy Height (NASA)☒ Land Cover (USGS)☐ Intact Forest Landscape (IntactForests)☒ WorldClim v.2 (WorldClim)

When selected from layers indexed previously, go through the process of checking what layers you specifically have.

**CUSTOM ENVIRONMENTAL DATA:**

ADD ENVIRONMENT DATA

REMOVE ENVIRONMENT DATA

**Environment Data 1:**

Environmental Data 1 Name: \*

Please provide the name of Environmental Data 1. Some example environmental data names might include "soil chemistry", "rainfall", "average temperature", etc.

Environmental Data 1 Description: \*

Please provide a short description of Environmental Data 1.

Environmental Data 1 Units: \*

Please provide the units of Environmental Data 1.

Environmental Data 1 Value: \*

Please provide the value of Environmental Data 1.





## CHAPTER 3

---

### Features

---

Coming soon!



1. TPPSc requires the following modules:

- [TPPS](#)

2. Inside your /sites/all/modules directory, download TPPSc by running:

```
git clone https://gitlab.com/TreeGenes/TPPSc.git
```

3. Then, enable TPPSc by running:

```
drush en tppsc
```

4. Finally, customize TPPSc to your liking using the TPPS settings form at <https://<your site>/admin/config/content/tpps>.

For more instructions, please see the [Drupal documentation page](#).