
Tripal Plant PopGen Submit Documentation

Peter Richter

Feb 14, 2023

Table of Contents:

1	Introduction	1
2	Features	3
2.1	Data Types and Standards	3
2.2	Data Accessibility	3
2.3	User Friendliness	4
2.4	Administrative Features	4
3	Administrators	5
3.1	Installation	5
3.2	Configuration	5
3.3	Login and Profile	6
3.4	Publication and Species Interface	8
3.5	Study Design	11
3.6	Plant Accession	18
3.7	Genotype, Phenotype, and Environment	21
3.8	Submission Summary	27
3.9	Input Validation	29
3.10	Data Submission	29
3.11	Submission Management	29
4	Users	31
4.1	Publication and Species	31
4.2	Experimental Conditions	34
4.3	Plant Accession	37
4.4	Genotype, Phenotype, and Environment	40
4.5	Submission Summary	47
5	Developers	49
5.1	Coding Style Specifications	49
5.2	Project Structure	49
5.3	File Manipulation	51
5.4	Where Does Data Go?	55

In the past 5 years alone, thousands of papers have been published on association genetics and/or landscape genomics of plants. Very little of this data is formally collected as georeferenced accessions with full integration of genotype and phenotype. TreeGenes has developed the Tripal Plant Pop-Gen Submit pipeline (TPPS), an open-source [Drupal](#) module built to extend the functionality of the [Tripal](#) toolset, to specifically capture data and metadata describing genotype, phenotype and environmental studies associated with landscape genomics or association genetics investigations. The workflow relies on a series of questions to properly describe the experimental design, including location, replication and treatments. These questions also guide the system on the types of raw and intermediate data to request. Raw sequence data and reference genomes are sent to the primary repositories and linked back to TreeGenes via NCBI/EBI accession numbers. Intermediate deliverables, such as assemblies and genotypes, are accepted within the TPPS workflow. TPPS is able to accommodate a wide range of designs common to forest genetics studies: landscape sampling, breeding plots, common gardens and growth chamber experiments.

TPPS enforces minimal reporting standards and associated biological ontologies to provide reusable data. The module employs standards established as the Minimal Information About a Plant Phenotyping Experiment ([MIAPPE](#)) to guide the collection of phenotypic data as well as the overall experimental design. The MIAPPE standards were developed from the objectives of transPLANT, European Plant Phenotyping Network and ELIXIR-EXCELERATE projects with the goal of developing reporting requirements to describe plant phenotyping experiments. MIAPPE integrates its minimal reporting standards with existing ontological frameworks. TreeGenes has implemented five of these: Plant Ontology (PO), Chemical Entities of Biological Interest ([ChEBI](#)), Trait Ontology (TO), Crop Ontology (CO), Phenotype And Trait Ontology ([PATO](#)), and a custom TreeGenes ontology that serves to hold traits in transition to established ontologies. Plant structure, development and trait terms are integrated with PATO and supporting ontologies via the Planteome project which enables comparative biology across the omics.

In addition to the traits and their associated plant structures, genotypic values are collected through TPPS. Currently, TPPS can accommodate both SNPs and microsatellites (SSRs), as well as other user-defined marker types. This marker data is collected in the context of the sequencing design, which may include genotyping assays, genotype-by-sequencing (GBS) approaches, transcriptomic AQ4 sequencing and whole genome resequencing. Community standards, such as the Variant Common Format, are preferred. However, alternatives consistent with minimal reporting are accepted and will be converted and stored for re-distribution in standard file formats. For all submissions with genotypic values, TPPS strongly encourages the user to reference a genome and version or provide an intermediate assembly (transcriptomic or genomic) from which the SNP calls were derived. In a final, optional step, environmental data for the georeferenced plants can be loaded directly from the layers used or as independent measurements conducted by the investigators. Following acceptance and validation of the data in the TPPS module, data is organized

and submitted to a database implementing the [CHADO](#) database schema, and an accession number is supplied to the user that provides a long-term reference to the entire dataset. TreeGenes works closely with journals focused on plant genetics to encourage researchers to submit these data at the time of publication. Accepted studies are available in TreeGenes under the ‘Tripal Plant Pop-Gen Submissions’ page, where users can download the associated flat files, organized by content type.

TPPS has many features that make data collection easier for administrators. Here are a few notable ones:

2.1 Data Types and Standards

- Support for genotype, phenotype, and environmental data and metadata
- Support for population, association, and landscape genetics studies
- Support for ontology standards, including the Minimum Information About a Plant Phenotyping Experiment ([MIAPPE](#))
- Support for standard genotyping file formats, such as .VCF
- Automatically submits data according to the Tripal CHADO database schema

2.2 Data Accessibility

- Data is standardized and stored in the local database so that other tools, for example, [CartograPlant](#), can easily collect and analyze it
- Restricted access to approved users of the site
- Accepted studies are associated and stored in the database with longterm accessions that can be used in publication
- The studies can be queried or downloaded (flatfiles) through the Tripal interface
- Display both complete and incomplete submissions on 'TPPS Submissions' user profile tab

2.3 User Friendliness

- Map thumbnails for quick visual validation
- Auto-complete appropriate fields based on information from the user profile
- Load data from [NCBI](#) based on a provided BioProject accession number
- Automatically parse file contents for submission to the CHADO schema
- Save user progress on incomplete submissions
- Form flexibility to ensure only the minimum necessary information is being required, but users may provide additional information if they choose

2.4 Administrative Features

- Administrator panel to manually approve completed submissions
- Configuration page to specify file upload locations, TPPS Admin email, etc.

3.1 Installation

1. TPPS requires the following modules:

- Tripal (v3.x)
- Tripal Chado
- Ultimate Cron
- JQuery Update
- Tripal EUtils

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

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

3. Then, enable TPPS by running:

```
drush en tpps
```

4. Finally, customize TPPS to your liking by visiting <https://<your site>/admin/config/content/tpps> and modifying the TPPS settings.

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

3.2 Configuration

There are some elements of TPPS that can be configured to fit your specific needs! After installing TPPS, log into your site as an administrator. Then, from your dashboard, go to Modules → TPPS → Configure, or go to `<base url>/admin/config/content/tpps`. There are a few settings you can customize as an administrator:

- TPPS Admin Email Address: the email address used to send administrative TPPS messages, such as notifications about submission status.

- Use environmental layers from CartograPlant: If CartograPlant is installed, TPPS can add an optional field to the environment section for environment layers, using the layer data pulled in through CartograPlant.
- If the Use environmental layers option is selected, you will be asked to identify which of the layer groups provided by CartograPlant contain environmental data that is relevant to TPPS. These groups will be used to decide which layers to present as environmental layer options to users in TPPS.
- TPPS Genotype Max Group: the maximum number of genotype records TPPS is allowed to try to submit together. Higher max group numbers will mean faster genotype file parsing jobs, but are more likely to cause errors with the Tripal Job daemon.
- Reference Genome Directory: the location of local reference genomes on your server. If left blank, TPPS will skip searching for local reference genomes.
- File upload locations:
 - Author files: the location to store secondary author files
 - Plant Accession files: the location to store plant accession files
 - Genotype files: the location to store .VCF, .FASTA, and genotype assay files
 - Phenotype files: the location to store phenotype data and metadata files

3.3 Login and Profile

Users can only access the TPPS form while they are logged into their Tripal account. This is so that a user can pick up where they left off, should they need to make their submission over multiple sessions. For example, if a user is submitting data through TPPS, then realizes that they need to collect some additional information before completing their submission, they can leave the TreeGenes site to collect their additional information, and when they return, all of the data that user entered previously will be stored on TreeGenes, and the user will not need to fill out all of the form fields again.

Checking if the user is logged in, loading relevant submissions, and redirecting or displaying the correct form is handled in the following snippet:

```
function tpps_main($form, &$form_state){

  global $user;
  $params = drupal_get_query_parameters();

  if (!isset($user->mail) and !isset($form_state['stage'])) {
    //first access, not logged in
    $form_state['stage'] = 'frontpage';

    front_create_form($form, $form_state);

    drupal_add_js(drupal_get_path('module', 'tpps') . "/js/tpps.js");

    return $form;
  }
  elseif (!isset($user->mail)) {
    //not first access, but not logged in
    $destination = drupal_get_destination();
    drupal_goto('user/login', array('query' => $destination));
  }
  elseif (isset($params['accession']) and !isset($form_state['stage'])) {
    //logged in, and with accession number in link
```

(continues on next page)

(continued from previous page)

```

    $old_form_state = variable_get('tpps_incomplete_' . $user->mail . $params[
    ↪ 'accession']);
    $form_state['saved_values'] = $old_form_state['saved_values'];
    $form_state['stage'] = $old_form_state['stage'];
    $form_state['accession'] = $old_form_state['accession'];
    $form_state['dbxref_id'] = $old_form_state['dbxref_id'];
  }
  else {
    //logged in, with plain link
    if (!isset($form_state['stage'])) {
      //first access
      $form_state['stage'] = 'frontpage';

      front_create_form($form, $form_state);

      drupal_add_js(drupal_get_path('module', 'tpps') . "/js/tpps.js");

      return $form;
    }
    else {
      //not first access
      if (!isset($form_state['saved_values']['frontpage']['accession']) or
    ↪ (isset($form_state['saved_values']['frontpage']['accession']) and $form_state[
    ↪ 'saved_values']['frontpage']['accession'] === 'new')) {
        //no submissions, or selected to create new submission
        tpps_init_project($form, $form_state);
        variable_set('tpps_incomplete_' . $user->mail . $form_state['saved_
    ↪ values']['frontpage']['accession'], $form_state);
      }
      $old_form_state = variable_get('tpps_incomplete_' . $user->mail . $form_
    ↪ state['saved_values']['frontpage']['accession']);
      $form_state['saved_values'] = $old_form_state['saved_values'];
      $form_state['stage'] = $old_form_state['stage'];
      $form_state['accession'] = $old_form_state['accession'];
      $form_state['dbxref_id'] = $old_form_state['dbxref_id'];
    }
  }

  // $form_state was loaded from saved variable, continue defining the form based on
  ↪ $form_state
  ...
}

```

If the user is not logged in, they are redirected to the login page, where they can login as an existing user or sign up to create a new account.

If the user is logged in, and they have any incomplete TPPS Submissions, they will be given the option to continue with one of their old submissions, or create a new TPPS Submission.

A screenshot of the landing page for a logged in user with a submission in progress can be seen below

Welcome to TPPS!

The Tripal PopGen Submit (TPPS) workflow provides researchers with a streamlined submission interface for studies resulting from any combination of genotype, phenotype, and environmental data for georeferenced forest trees. TPPS will guide users through questions about their study design and data in order to collect information on trees, genotypes, and phenotypes if applicable. Phenotypic, genotypic, and environmental descriptors will be mapped to ontologies where possible and the collected metadata will enable this information to be displayed in CartograTree. An accession number will be provided to the user following successful completion that uniquely identifies this study in the database. This number should be used in the manuscript describing this work. Specific flat-files and metadata associated with this identifier can be accessed by the public with this information.

To get started, you will need to have a few things handy:

- An enabled and approved TreeGenes account - you can create one [here](#). There may be a waiting period to have your account approved by a TreeGenes administrator.
- Information about the paper connected to your study, and the organisms being studied. This must include at least the following:
 - Primary author of the publication
 - Organization of the Primary Author
 - Publication year
 - Publication title
 - Publication abstract
 - Journal the publication can be found in
 - Genus and species of the organism(s) being studied
- Metadata about the study itself. This must include at least the following:
 - Dates when the study took place (month, year)
 - The location of the study (coordinates or country/region)
 - Type of study (natural population, growth chamber, common garden, etc.)
 - Type of data collected (some combination of genotype, phenotype, and environmental)
 - Relevant quantitative information based on the type of study, such as the seasons of a natural population study, or the soil pH of a growth chamber study
- Geographic locations of the trees (for common garden studies, this would be the location of the source tree). This should be a spreadsheet with a column for tree identifiers and column(s) for the location of each tree.
- Genotypic and/or phenotypic data and metadata (depending on the type(s) of data collected).

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

Would you like to load an old TPPS submission, or create a new one? [Create new TPPS Submission](#)

[CONTINUE TO TPPS](#)

3.4 Publication and Species Interface

The first set of fields in TPPS is the publication and species interface, where users upload data about their publication and up to 5 studied species. The form fields and their properties are as follows:

- User Info: `fieldset`
 - Primary Author: `textfield` - auto-populates with the name registered to their Tripal account. If the primary author is changed, autocomplete options are provided from the `chado.contact` table.
 - Organization: `textfield` - auto-populates with the organization registered to their Tripal account. If the organization is changed, autocomplete options are provided from the `chado.contact` table.

- Publication: fieldset
 - Secondary Authors: fieldset
 - * Secondary Author x: textfield - autocomplete options from the chado.contact table
 - * >30 Secondary Authors: checkbox
 - * Secondary Authors file: managed_file - spreadsheet of secondary authors. This field is only visible if the '>30 Secondary Authors' checkbox is checked.
 - Publication status: select - options 'In Preparation or Submitted', 'In press', and 'Published'
 - Publication Year: select - options '1990' to '2018'
 - Publication Title: textfield
 - Publication Abstract: textarea
 - Publication Journal: textfield - autocomplete options from chado.pub table
- Plant Species: fieldset
 - Species x: textfield - autocomplete options from the chado.organism table
 - Up to 5 different species are allowed per submission

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

Primary Author: *

Organization: *

▼ PUBLICATION INFORMATION:

▼ SECONDARY AUTHOR INFORMATION

ADD SECONDARY AUTHOR REMOVE SECONDARY AUTHOR

Publication Status: *

Title of Publication/Study: *

Abstract/Description: *

Journal: *

▼ STUDY COVER PHOTO: (OPTIONAL)

Please upload a cover photo for your study. This photo will be displayed at the top of the landing page of the study.

Choose File No file chosen

UPLOAD

▼ ORGANISM INFORMATION:

ADD ORGANISM REMOVE ORGANISM

Species 1

☒ This species is a tree.

3.5 Study Design

The second set of fields in TPPS is the Study Design section, where users upload metadata about their experiment. The form fields and their properties are as follows:

- Study Start Date: `fieldset`
 - Start Date Year: `select` - options '1970' to '2018'
 - Start Date Month: `select` - options 'January' through 'December'
- Study End Date: `fieldset`
 - End Date Year: `select` - options '1970' to '2018'
 - End Date Month: `select` - options 'January' through 'December'
- Study Location: `fieldset`
 - Coordinate Projection: `select` - options 'WGS 84', 'NAD 83', 'ETRS 89', 'Custom Location (street address)'
 - Coordinates: `textfield` - only visible if the user selects 'WGS 84', 'NAD 83', or 'ETRS 89' from 'Coordinate Projection'
 - Custom Location: `textfield` - only visible if the user selects 'Custom Location' from 'Coordinate Projection'
 - Map Thumbnail: Google Maps iframe that allows the user to visually validate the location they have provided
- Data Type: `select` - options 'Genotype', 'Genotype x Phenotype', 'Genotype x Environment', 'Genotype x Phenotype x Environment', 'Phenotype x Environment'
- Study Type: `select` - options 'Natural Population (Landscape)', 'Growth Chamber', 'Greenhouse', 'Experimental/Common Garden', 'Plantation'

A screenshot of the Study Date and Study Location fields can be seen below:

STUDY DESIGN

Data Type: *

Genotype x Phenotype (and/or manual environmental data) x Environmental layers

Study Type: *

Experimental/Common Garden

COMMON GARDEN INFORMATION:

Irrigation Type: *

Irrigation from bottom

Salinity controlled or uncontrolled: *

Uncontrolled

Average Salinity Value: *

Biotic Environment: *

☐ Herbivores
 ☐ Mutualists
 ☒ Pathogens
 ☐ Endophytes
 ☐ Other
 ☐ None

Seasons: *

☐ Spring
 ☒ Summer
 ☐ Fall
 ☐ Winter

If you do not know which season your samples were collected, please select all.

3.5.1 Natural Population

- Natural Population: `fieldset` - only visible if the user selects 'Natural Population' from 'Study Type'
 - Season: `checkboxes` - options 'Spring', 'Summer', 'Fall', 'Winter'
 - Assessments: `select` - options '1' to '30'

A screenshot of the Natural Population fields can be seen below:

Data Type: *

Study Type: *

▼ NATURAL POPULATION/LANDSCAPE INFORMATION:

Seasons (select all that apply): *

☒ Spring

☒ Summer

☐ Fall

☐ Winter

Number of times the populations were assessed (on average): *

3.5.2 Growth Chamber

- Growth Chamber: fieldset - only visible if the user selects 'Growth Chamber' from 'Study Type'
 - CO2 Info: fieldset
 - * CO2 Control: select - options 'controlled', 'uncontrolled'
 - * CO2 Value: textfield
 - Humidity Info: fieldset
 - * Humidity Control: select - options 'controlled', 'uncontrolled'
 - * Humidity Value: textfield
 - Light Intensity Info: fieldset
 - * Light Intensity Control: select - options 'controlled', 'uncontrolled'
 - * Light Intensity Value: textfield
 - Temperature Info: fieldset
 - * Average High Temperature: textfield
 - * Average Low Temperature: textfield
 - Rooting Info: fieldset
 - * Rooting Type: select - options 'Aeroponics', 'Hydroponics', and 'Soil'
 - * Soil: fieldset - only visible if the user selects 'Soil' from 'Rooting Type'
 - Soil Type: select - options 'Sand', 'Peat', 'Clay', 'Mixed', 'Other'
 - Custom Soil Type: textfield - only visible if the user selects 'Other' from 'Soil Type'
 - Soil Container Type: textfield
 - * PH: fieldset
 - PH Control: select - options 'controlled', 'uncontrolled'
 - PH Value: textfield
 - * Treatments: checkboxes - options 'Seasonal Environment', 'Air temperature regime', 'Soil Temperature regime', 'Antibiotic regime', 'Chemical administration', 'Disease status', 'Fertilizer regime', 'Fungicide regime', 'Gaseous regime', 'Gravity Growth hormone regime', 'Mechanical treatment'

‘Mineral nutrient regime’, ‘Humidity regime’, ‘Non-mineral nutrient regime’, ‘Radiation (light, UV-B, X-ray) regime’, ‘Rainfall regime’, ‘Salt regime’, ‘Watering regime’, ‘Water temperature regime’, ‘Pesticide regime’, ‘pH regime’, ‘other perturbation’

- * Treatments Description: `textfield` - if an option from ‘Treatments’ is selected, users must provide a description of the treatment

A screenshot of the Growth Chamber fields can be seen below:

Data Type: * Genotype x Phenotype x Environment
Study Type: * Growth Chamber

GROWTH CHAMBER INFORMATION:

CO2 controlled or uncontrolled: * Controlled

Controlled CO2 Value: * 10

Air Humidity controlled or uncontrolled: * Uncontrolled

Average Air Humidity Value: * 45%

Light Intensity controlled or uncontrolled: * Uncontrolled

Average Light Intensity Value: * 200

TEMPERATURE INFORMATION:

Please provide temperatures in Degrees Celsius

Average High Temperature: * 40

Average Low Temperature: * 20

ROOTING INFORMATION:

Rooting Type: * Hydroponics

pH controlled or uncontrolled: * Uncontrolled

Average pH Value: * 7.5

TREATMENTS: *

☒ Seasonal Environment

☐ Air temperature regime

☐ Soil Temperature regime

☐ Antibiotic regime

☐ Chemical administration

☐ Disease status

☐ Fertilizer regime

☐ Fungicide regime

☐ Gaseous regime

☐ Gravity Growth hormone regime

3.5.3 Greenhouse

- Greenhouse: `fieldset` - only visible if the user selects 'Greenhouse' from 'Study Type'
 - Humidity Info: `fieldset`
 - * Humidity Control: `select` - options 'controlled', 'uncontrolled'
 - * Humidity Value: `textfield` - only visible if the user selects 'controlled' from 'Humidity Control'
 - Light Intensity Info: `fieldset`
 - * Light Intensity Control: `select` - options 'controlled', 'uncontrolled'
 - * Light Intensity Value: `textfield` - only visible if the user selects 'controlled' from 'Light Intensity Control'
 - Temperature Info: `fieldset`
 - * Average High Temperature: `textfield`
 - * Average Low Temperature: `textfield`
 - Rooting Info: `fieldset`
 - * Soil: `fieldset`
 - Soil Type: `select` - options 'Sand', 'Peat', 'Clay', 'Mixed', 'Other'
 - Custom Soil Type: `textfield` - only visible if the user selects 'Other' from 'Soil Type'
 - Soil Container Type: `textfield`
 - * PH: `fieldset`
 - PH Control: `select` - options 'controlled', 'uncontrolled'
 - PH Value: `textfield` - only visible if the user selects 'controlled' from 'PH Control'
 - * Treatments: `checkboxes` - options 'Seasonal Environment', 'Air temperature regime', 'Soil Temperature regime', 'Antibiotic regime', 'Chemical administration', 'Disease status', 'Fertilizer regime', 'Fungicide regime', 'Gaseous regime', 'Gravity Growth hormone regime', 'Mechanical treatment', 'Mineral nutrient regime', 'Humidity regime', 'Non-mineral nutrient regime', 'Radiation (light, UV-B, X-ray) regime', 'Rainfall regime', 'Salt regime', 'Watering regime', 'Water temperature regime', 'Pesticide regime', 'pH regime', 'other perturbation'
 - * Treatments Description: `textfield` - if an option from 'Treatments' is selected, users must provide a description of the treatment

A screenshot of the Greenhouse fields can be seen below:

Data Type: *

Study Type: *

▼ **GREENHOUSE INFORMATION:**

Air Humidity controlled or uncontrolled: *

Light Intensity controlled or uncontrolled: *

TEMPERATURE INFORMATION:

Please provide temperatures in Degrees Celsius

Average High Temperature: *

Average Low Temperature: *

ROOTING INFORMATION:

Rooting Type: *

pH controlled or uncontrolled: *

Controlled pH Value: *

TREATMENTS: *

- ☐ Seasonal Environment
- ☐ Air temperature regime
- ☐ Soil Temperature regime
- ☐ Antibiotic regime
- ☐ Chemical administration
- ☐ Disease status
- ☐ Fertilizer regime
- ☐ Fungicide regime
- ☐ Gaseous regime
- ☐ Gravity Growth hormone regime
- ☐ Mechanical treatment
- ☐ Mineral nutrient regime
- ☐ Humidity regime
- ☐ Non-mineral nutrient regime

3.5.4 Common Garden

- Common Garden: fieldset - only visible if the user selects 'Experimental/Common Garden' from 'Study Type'
 - Irrigation: fieldset
 - * Irrigation Type: select - options 'Irrigation from top', 'Irrigation from bottom', 'Drip Irrigation', 'Other', 'No Irrigation'
 - * Custom Irrigation Type: textfield - only visible if the user selects 'Other' from 'Irrigation Type'

- Salinity Info: `fieldset`
 - * Salinity Control: `select` - options 'controlled', 'uncontrolled'
 - * Salinity Value: `textfield`
- Biotic Environment Info: `fieldset`
 - * Biotic Environment Type: `select` - options 'Herbivores', 'Mutulists', 'Pathogens', 'Endophytes', 'Other', 'None'
 - * Custom Biotic Environment Type: `textfield` - only visible if the user selects 'Other' from 'Biotic Environment Type'
- Season: `checkboxes` - options 'Spring', 'Summer', 'Fall', 'Winter'
- Treatments: `checkboxes` - options 'Seasonal environment', 'Antibiotic regime', 'Chemical administration', 'Disease status', 'Fertilizer regime', 'Fungicide regime', 'Gaseous regime', 'Gravity Growth hormone regime', 'Herbicide regime', 'Mechanical treatment', 'Mineral nutrient regime', 'Non-mineral nutrient regime', 'Salt regime', 'Watering regime', 'Pesticide regime', 'pH regime', 'Other perturbation'
- Treatments Description: `textfield` - if an option from 'Treatments' is selected, users must provide a description of the treatment

A screenshot of the Common Garden fields can be seen below:

Data Type: * Genotype x Phenotype x Environment
Study Type: * Experimental/Common Garden

COMMON GARDEN INFORMATION:

Irrigation Type: * No Irrigation

Salinity controlled or uncontrolled: * Controlled

Controlled Salinity Value: *

Biotic Environment: *

☒ Herbivores
☒ Mutulists
☐ Pathogens
☒ Endophytes
☐ Other
☐ None

Seasons: *

☒ Spring
☒ Summer
☐ Fall
☐ Winter

TREATMENTS:

☐ My Common Garden experiment used treatments/regimes/perturbations.

3.5.5 Plantation

- Plantation: `fieldset` - only visible if the user selects 'Plantation' from 'Study Type'
 - Season: `checkboxes` - options 'Spring', 'Summer', 'Fall', 'Winter'
 - Assessments: `textfield`
 - Treatments: `checkboxes` - options 'Seasonal environment', 'Antibiotic regime', 'Chemical administration', 'Disease status', 'Fertilizer regime', 'Fungicide regime', 'Gaseous regime', 'Gravity Growth hormone regime', 'Herbicide regime', 'Mechanical treatment', 'Mineral nutrient regime', 'Non-mineral nutrient regime', 'Salt regime', 'Watering regime', 'Pesticide regime', 'pH regime', 'Other perturbation'
 - Treatments Description: `textfield` - if an option from 'Treatments' is selected, users must provide a description of the treatment

A screenshot of the Plantation fields can be seen below:

3.6 Plant Accession

The third set of fields in TPPS is the Plant Accession section, where users upload information about where plants are located. The form fields and their properties are as follows:

- Plant Accession: `fieldset` - only visible if number of species in [Publication and Species Interface](#) is 1, or if the 'Separate Plant Accession' checkbox is unchecked.
 - Plant Accession File: `managed_file` - spreadsheet of plant locations
 - Plant Accession Columns: `fieldset` of `select` elements - user will define which of their columns contain the Plant ID, and which of their columns contain location information. If the number of species is >1, user will also define which of their columns contain the genus and species of each plant.
 - Coordinate Projection: `select` - options 'WGS 84', 'NAD 83', 'ETRS 89'. Only visible if 'Latitude' or 'Longitude' columns are selected in 'Plant Accession Columns'
 - Map Thumbnail: Google Maps iframe that allows the user to visually validate the plant locations they have provided
- Separate Plant Accession: `checkbox`

- Separate Plant Accession Information: `fieldset` - only visible if the 'Separate Plant Accession' checkbox is checked
 - Plant Accession File **x**: `managed_file` - spreadsheet of plant locations of species **x**
 - Plant Accession Columns **x**: `fieldset` of `select` elements - user will define which of their columns contain the Plant ID, and which of their columns contain location information

Allowed file extensions: txt csv xlsx

 tpps_accession_large.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: Plant Identifier and Location. If your plants are located based on a population group, you can provide the population group column and a mapping of population group to location below.

tree identif	<input type="text" value="Plant Identifier"/>	latitude	<input type="text" value="Latitude"/>	longitude	<input type="text" value="Longitude"/>
	TPPS-DELETE-ME-100		42		72
	TPPS-DELETE-ME-101		42		70
	TPPS-DELETE-ME-102		44		71

Please note that if you remain idle on a TPPS page with files for more than 6 hours, you run the risk of having those files reset if you do not click "Save".

☐ My file has no header row

Coordinate Projection



☒ The provided GPS coordinates are exact

3.7 Genotype, Phenotype, and Environment

The fourth set of fields in TPPS is the Genotype, Phenotype, and Environment section, where users upload Genotypic, Phenotypic, and Environmental data and metadata about each organism. In multi-species submissions, users can reuse data from the previous species by checking the '<Genotype/Phenotype/Environmental> information for <organism i> is the same as <genotype/phenotype/environmental> information for <organism i-1>.' box, which is only available after the first organism. The form fields and their properties are as follows:

- Species x: `fieldset`

3.7.1 Phenotype

- Phenotype Information: `fieldset` - only visible if the user selects 'Genotype x Phenotype', 'Genotype x Phenotype x Environment', or 'Phenotype x Environment' from 'Data Type' in [Study Design](#)
 - Phenotype x: `fieldset`
 - * Phenotype Name: `textfield` - autocomplete options from `chado.phenotype` table
 - * Phenotype Attribute: `textfield` - autocomplete options from `chado.phenotype` table
 - * Phenotype Description: `textarea`
 - * Phenotype Units: `textfield` - autocomplete options from `chado.phenotypeprop` table
 - * Phenotype Structure: `textfield` - autocomplete options from `chado.phenotype` table - only visible if the 'Phenotype x has a structure descriptor' checkbox is checked
 - * Phenotype Value Range : `textfield` - autocomplete options from `chado.phenotypeprop` table - only visible if the 'Phenotype x has a value range' checkbox is checked
 - Phenotype Metadata file: `managed_file` - spreadsheet of metadata about each phenotype - only visible if the 'I would like to upload a phenotype metadata file' checkbox is checked
 - Phenotype Metadata File Columns: `fieldset` of `select` elements - user will define which of their columns contain the Phenotype Name/Identifier, Phenotype Attribute, Phenotype Description, Phenotype Units, Phenotype Structure, Max/Min Phenotype Values - only visible if the 'I would like to upload a phenotype metadata file' checkbox is checked
 - Phenotype File: `managed_file` - spreadsheet of phenotypes
 - Phenotype File Columns: `fieldset` of `select` elements - user will define which of their columns contain the Plant ID, Phenotype Name/Identifier, and Phenotype value

A screenshot of the manual phenotype information fields can be seen below:

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: *

Phenotype 1 Description: *

Please provide a short description of Phenotype 1

Phenotype 1 Units: *

Phenotype 1 Structure: *

☐ Phenotype 1 has a value range

☐ Phenotype 1 is a binary phenotype

☐ Phenotype 1 is an environmental phenotype

A screenshot of the phenotype metadata file field can be seen below:

Phenotype Metadata File: Please upload a file containing columns with the name, attribute, structure, description, and units of each of your phenotypes: *

 tpps_phenotype_meta.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: Phenotype name

Phenotype Name	Phenotype Description	Phenotype Calculation
Phenotype Name/Identifier ▼	Description ▼	Attribute ▼
Units		
a2	Per-planted-area cross-sectional area of stems at breast height (1.3m)	Area/area
b2	The number of trees with broken arms	The number of applicable trees in each plot
c2	The number of trees with crooked stems	The number of applicable trees in each plot

Please note that if you remain idle on a TPPS page with files for more than 6 hours, you run the risk of having those files reset if you do not click "Save".

☐ My file has no header row

A screenshot of the phenotype data file field can be seen below:

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

 tpps_phenotype.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: Plant Identifier, Phenotype name, and Value(s)

Site ID	Tree ID	f2	n1	b2
<input type="text" value="N/A"/>	<input type="text" value="Plant Identifier"/>	<input type="text" value="Phenotype Data"/>	<input type="text" value="Phenotype Data"/>	<input type="text" value="Phenotype Data"/>
Site_N	N_6_1_1	0	32	1
Site_N	S_50_1_3	0	15	1
Site_N	N_15_1_4	5	31	2

Please note that if you remain idle on a TPPS page with files for more than 6 hours, you run the risk of having those files reset if you do not click "Save".

☐ My file has no header row

3.7.2 Genotype

- Genotype Information: `fieldset` - only visible if the user selects 'Genotype', 'Genotype x Phenotype', 'Genotype x Environment', or 'Genotype x Phenotype x Environment' from 'Data Type' in [Study Design](#)
 - Genotype Marker Type: `checkboxes` - options 'SNPs', 'SSRs/cpSSRs', 'Other'
 - Genotype SNPs: `fieldset` - only visible if the user selects 'SNPs' from 'Genotype Marker Type'
 - * SNPs Genotyping Design: `select` - options 'GBS', 'Targeted Capture', 'Whole Genome Resequencing', 'RNS-Seq', 'Genotyping Array'
 - * GBS Type: `select` - options 'RADSeq', 'ddRAD-Seq', 'NextRAD', 'RAPTURE', 'Other' - only visible if the user selects 'GBS' from 'SNPs Genotyping Design'
 - * GBS Custom Type: `textfield` - only visible if the user selects 'Other' from 'GBS Type'
 - * Targeted Capture Type: `select` - options 'Exome Capture', 'Other' - only visible if the user selects 'Targeted Capture' from 'SNPs Genotyping Design'
 - * Targeted Capture Custom Type: `textfield` - only visible if the user selects 'Other' from 'Targeted Capture Type'
 - Genotype SSRs/cpSSRs Type: `textfield` - only visible if the user selects 'SSRs/cpSSRs' from 'Genotype Marker Type'
 - Genotype Other Marker Type: `textfield` - only visible if the user selects 'Other' from 'Genotype Marker Type'
 - Reference Genome: `select` - stored reference genomes, as well as 'I can provide a URL to the website of my reference file(s)', 'I can provide a GenBank accession number (BioProject, WGS, TSA) and select

- assembly file(s) from a list', 'I can upload my own reference genome file', 'I can upload my own reference transcriptome file', 'I am unable to provide a reference assembly'
- BioProject Accession: `textfield` - only visible if the user selects 'I can provide a GenBank accession number (BioProject, WGS, TSA) and select assembly file(s) from a list' from 'Reference Genome'
 - * NCBI Assembly Accessions: `checkboxes` - options pulled directly from NCBI
- URL or Manual Assembly File: `fieldset` - Tripal FASTA Loader fields - only visible if the user selects 'I can provide a URL to the website of my reference file(s)', 'I can upload my own reference genome file', or 'I can upload my own reference transcriptome file' from 'Reference Genome'
- Genotype File Type: `checkboxes` - options 'Genotype Spreadsheet/Assay', 'Assay Design', 'VCF'. 'Assay Design' only visible if the user selects 'SNPs' from 'Genotype Marker Type'.
- Genotype VCF File: `managed_file` - .VCF file of genotypes - only visible if the user selects 'VCF' from 'Genotype File Type'
- Genotype File: `managed_file` - spreadsheet of genotypes - only visible if the user selects 'Genotype Spreadsheet/Assay' from 'Genotype File Type'
- Genotype File Columns: `fieldset` of select elements - user will define which of their columns contain the Plant ID and Genotype Data
- Assay Design File: `managed_file` - Assay design file - only visible if the user selects 'Assay Design' from 'Genotype File Type'

A screenshot of the genotype marker type fields can be seen below:

▸ GENOTYPE INFORMATION:

Marker Type (select all that apply): *

☒ SNPs
☐ SSRs/cpSSRs
☐ Indels
☐ Other

▸ SNPS INFORMATION:

Define Experimental Design: * GBS

GBS Type: * ddRAD-Seq

Reference Assembly used: * Carica papaya v0.4

A screenshot of the genotype reference fields can be seen below:

Reference Assembly used: *

I can upload my own reference genome file

Tripal FASTA Loader

FASTA Upload

Please provide the FASTA file. The file must have a .fasta extension.

Existing Files --Select a file--

You may select a file that is already uploaded.

Choose File No file chosen

UPLOAD

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".

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.

A screenshot of the genotype file fields can be seen below:

Genotype File Types (select all that apply): *

☒ SNPs Genotype Assay

☐ Assay Design

☐ SNPs Associations

☒ VCF

SNPs Genotype Assay File: please provide a spreadsheet with columns for the Plant ID of genotypes used in this study: *

Choose File No file chosen

UPLOAD

Please upload a spreadsheet file containing SNP Genotype Assay data. The format of this file is very important! The first column of your file should contain plant identifiers which match the plant identifiers you provided in your plant accession file, and all of the remaining columns should contain SNP data.

Genotype VCF File: *

Choose File No file chosen

UPLOAD

3.7.3 Environment

- Environment Information: `fieldset` - only visible if the user selects 'Environment', 'Phenotype x Environment', 'Genotype x Environment', or 'Genotype x Phenotype x Environment' from 'Data Type' in [Study Design](#)
 - CartograPlant Environmental Layers: `fieldset` - only visible if CartograPlant Layers are enabled in TPPS admin configuration and the 'I used environmental layers in my study that are indexed by CartograPlant.' checkbox is checked.
 - * CartograPlant Environmental Layer **x**: `checkbox` - Indicates if the CartograPlant Environmental Layer **x** was used.
 - CartograPlant Environmental Layer Parameters: `fieldset` - only visible if CartograPlant Layers are enabled in TPPS admin configuration and the 'I used environmental layers in my study that are indexed by CartograPlant.' checkbox is checked.
 - * CartograPlant Environmental Layer **x** Parameters: `checkboxes` - options of possible parameter types for the selected CartograPlant Environmental Layer. Each CartograPlant Environmental Layer **x** Parameters checkboxes set is only visible if that layer was selected in 'CartograPlant Environmental Layers'.
 - Custom Environmental Data: `fieldset` - only visible if the 'I have environmental data that I collected myself.' checkbox was checked.
 - * Environmental Data **x**: `fieldset`
 - Environmental Data Name: `textfield`
 - Environmental Data Description: `textfield`
 - Environmental Data Units: `textfield`
 - Environmental Data Value: `textified`

3.8 Submission Summary

Once all of the data on the previous pages has been submitted, users will be brought to a summary page where they can view a brief overview of the data they have provided so far. A sample screenshot of the summary can be seen below.

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

Publication Title	TPPS Example Submission
Primary Author	Peter Richter
Organization	University of Connecticut
Publication Status	In Press
Publication Abstract	Abstract placeholder for Example TPPS Submission.
Publication Journal	Journal of bioscience and bioengineering
Species 1	Pinus taeda
Experiment Starting Date	July 1973
Experiment Ending Date	June 1984
Study Location	41.8077° N, 72.2540° W
Data Type	Genotype x Phenotype x Environment
Study Type	Natural Population (Landscape)
Seasons Assessed	Spring Summer
Average number of assessments	15
Tree Accession File	tpps_accession_test.xlsx
Pinus taeda Phenotype Metadata File	phenotype metadata.xlsx
Pinus taeda Phenotype File	phenotype_file.xlsx
Pinus taeda Genotype Marker Types	SNPs
Pinus taeda Genotyping Design	GBS
Pinus taeda GBS Type	ddRAD-Seq
Pinus taeda Uploaded Reference File	661
Pinus taeda Genotype VCF File	test.vcf
Cartogratree Layer: Minimum temperature January Parameters	Minimum temperature
Cartogratree Layer: Minimum temperature February Parameters	Minimum temperature

Users can also add additional comments or supplemental files to their submission if they feel that these are necessary.

If you have any additional comments about this submission you would like to include, please write them here:

▼ ADDITIONAL FILES

If there are any additional files you would like to include with your submission, please upload up to 10 files here.

ADD FILE

REMOVE FILE

Supplemental File 1

Choose File No file chosen

UPLOAD

Supplemental File 2

Choose File No file chosen

UPLOAD

3.9 Input Validation

Once the 4 sets of fields have been filled out by the user, their data is validated to ensure completion and integrity.

- All fields are required when they are visible, and are therefore checked for completeness.
- Plant IDs from Phenotype and Genotype files are checked against Plant IDs in Plant Accession files to ensure there is not any data without plants.
- Scaffold/Chromosome IDs from .VCF files or Genotype assay files are checked against scaffold/chromosome IDs in assembly files to ensure there are no scaffolds without position.
- Phenotype names/IDs from phenotype data files are checked against phenotype names/IDs in phenotype meta-data files to ensure there are no undefined phenotypes.
- Users are allowed to upload plants in Plant Accession files without Genotype/Phenotype data, but not Genotype/Phenotype data without plant locations.

Input validation is broken up into steps, once after each set of fields. This is so that the user cannot continue to the next set of fields if they have incomplete or invalid data in their current set, and so that a user does not need to go back to previous sets of fields to correct data later.

3.10 Data Submission

Once the Input validation is completed for each of the 4 sets of fields, the data is saved in a persistent variable in the database, where it will wait until it is approved or rejected by an administrator. Both the user and administrator will be alerted once the submission has been completed, and again when the submission has been approved or rejected.

Upon approval, the data from the persistent variable is parsed, organized, and submitted according to the CHADO schema. TPPS makes sure not to overwrite existing entries in CHADO.

Persistent variables that TPPS creates will be removed from the database upon uninstallation.

3.11 Submission Management

Once a submission is completed, it is not available to the public until it has been approved by an administrator. This can be done from the TPPS admin approval panel, which can be accessed in one of two ways:

1. Click the link in the alert sent to the site admin email
2. Navigate to `http://<site domain>/tpps-admin-panel/<accession number>`

Once you are on the submission approval panel, you can either fill out the rejection reason field and click **Reject** to reject the submission, or you can check the **This submission has been reviewed and approved** box, and click **Approve** to approve it.

If the submission is approved the user will receive a notification via email, and the data that was not added through file uploads will be added to chado. The remaining data, which was added through file uploads, will be added to chado later during a tripal job. This data can take a long time to add, which is why it is added in a tripal job rather than immediately after approval.

If the submission is rejected, the user will receive a notification via email, and the submission will move from the completed state back to an incomplete submission. That way, the user can see the comments from the administrator, and can make appropriate changes to their submission.

3.11.1 Managing Submission Tags

One of the newer features of TPPS is the Submission Tagging system. Submission tags appear on the TPPS details page, as well as the TPPS admin panel. The TPPS, TPPSc, and data type tags are all added to submissions automatically. However, you can manually add and remove custom tags for a submission through the TPPS admin panel:

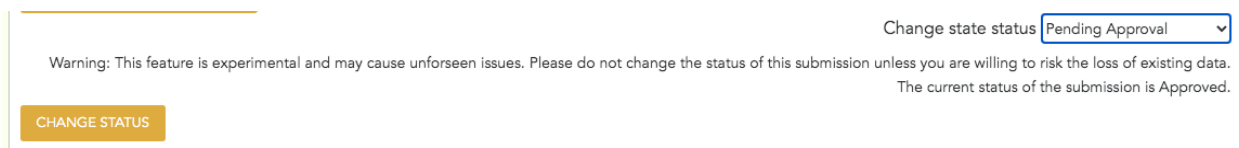
1. Navigate to `http://<site domain>/tpps-admin-panel/<accession number>#tags`
2. Click the tags in the “Available Tags” section that you would like to add.
3. Click the remove button on tags in the “Current Tags” section that you would like to remove.
4. To create a new tag or change the name or color of an existing tag, click the “Manage TPPS Submission Tags” link.

Tags can be added and removed from submissions regardless of their status, meaning you can manage the tags of submissions anywhere in the submission process, from *Incomplete* to *Approved*.

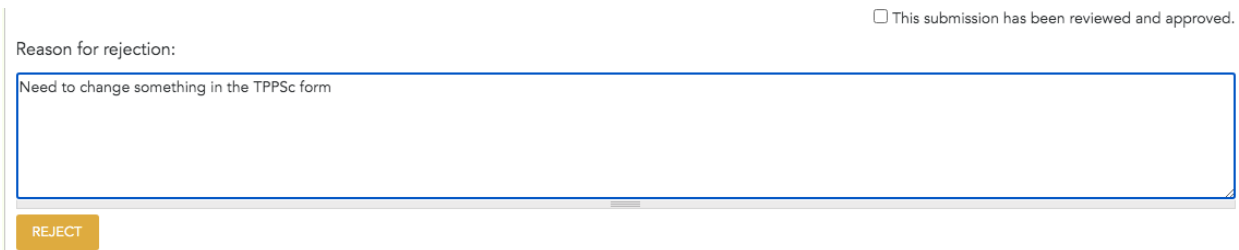
3.11.2 Changing a Submission State

Sometimes you may want to go back and change parts of a completed submission. However, if you try to access the TPPS or TPPSc form for a completed submission, you should get an “Access denied” message. To access the form for a completed submission, you must first change the state of the submission back to *Incomplete*. **Please note:** this process will add the “Under construction” warning to the TPPS details page for the submission. To change the status of the submission back to *Incomplete*:

1. Navigate to `http://<site domain>/tpps-admin-panel/<accession number>`
2. Scroll down to the “Change State Status” field.
3. Select “Pending Approval” from the drop-down.



4. Click the “Change Status” button.
5. The page will now reload, and the submission will appear as it normally does when pending approval. Scroll back down to the “Reason for rejection” section and write a note about why you need to go back to the TPPS or TPPSc form for this submission.



6. Click the “Reject” button.
7. You should now be able to access the TPPS or TPPSc form for this submission.

4.1 Publication and Species

The first set of fields is the Author and Species Information section. Here you will be asked to provide information about the publication, its authors, and the species studied.

4.1.1 User Information

- Primary Author: text field - The name of the first author of your publication.
- Organization: text field - The name of the organization of the first author of your publication.

4.1.2 Publication

- Secondary Authors: Click the 'Add Secondary Author' or 'Remove Secondary Author' buttons to add or remove secondary authors, respectively.
- Publication Status: drop-down menu - The status of the publication in the publishing process: 'In Preparation or Submitted', 'In press', or 'Published'
- Publication Year: drop-down menu - The year the publication was published or submitted: years from 1990 to 2018
- Publication Title: text field - The title of the publication.
- Publication Abstract: text area - The abstract for the publication.
- Publication Journal: text field - The journal the publication was published in.
- Study Cover Photo: file upload - Optional cover photo for the study.

4.1.3 Species

- Species #: text field - The genus and species of the species studied.
- Species # is tree: checkbox - Indicate whether or not the species is a tree.

A screenshot of the Publication and Species page can be seen below:

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

Primary Author: *

Organization: *

PUBLICATION INFORMATION:

SECONDARY AUTHOR INFORMATION

ADD SECONDARY AUTHOR

REMOVE SECONDARY AUTHOR

Publication Status: *

- Select -

Title of Publication/Study: *

Abstract/Description: *

Journal: *

STUDY COVER PHOTO: (OPTIONAL)

Please upload a cover photo for your study. This photo will be displayed at the top of the landing page of the study.

Choose File

No file chosen

UPLOAD

ORGANISM INFORMATION:

ADD ORGANISM

REMOVE ORGANISM

Species 1

☒ This species is a tree.

4.2 Experimental Conditions

The second set of fields in TPPS is the Study Design section. Here you will be asked to provide metadata about the study, such as when and where it took place, the type of data collected, and environmental factors.

4.2.1 Experiment Dates

- Start Date: two drop-down menus - The month and year that the experiment started
- End Date: two drop-down menus - The month and year that the experiment ended

4.2.2 Data Type

- Data Type: drop-down menu - The type of data collected during the study: 'Genotype', 'Phenotype', 'Environment', 'Genotype x Environment', 'Phenotype x Environment', 'Genotype x Phenotype', 'Genotype x Phenotype x Environment'

4.2.3 Study Type

- Study Type: drop-down menu - The type of study: 'Natural Population', 'Growth Chamber', 'Greenhouse', 'Experimental/Common Garden', 'Plantation'

4.2.4 Study Details

Natural Population

Natural Population fields are only visible when 'Natural Population' was selected from the 'Study Type' field.

- Season: checkboxes - The season(s) that the plants were assessed.
- Assessments: drop-down menu - The number of times, on average, the plants were assessed.

Growth Chamber

Growth Chamber fields are only visible when 'Growth Chamber' was selected from the 'Study Type' field.

- CO2 Control: drop-down menu - Whether the CO2 level was controlled during the study: 'controlled', 'uncontrolled'.
- CO2 Value: text field - The CO2 level during the study. If the level was not controlled, then provide the average CO2 level.
- Humidity Control: drop-down menu - Whether the Air Humidity was controlled during the study: 'controlled', 'uncontrolled'
- Humidity Value: text field - The Air Humidity during the study. If the level was not controlled, then provide the average Air Humidity.
- Light Intensity Control: drop-down menu - Whether the Light Intensity was controlled during the study: 'controlled', 'uncontrolled'
- Light Intensity: text field - The Light Intensity during the study. If the level was not controlled, then provide the average Light Intensity.

- Average High Temperature: text field - The average high temperature.
- Average Low Temperature: text field - The average low temperature.
- Rooting Type: drop-down menu - The type of rooting: 'Aeroponics', 'Hydroponics', 'Soil'
- Soil fields are only visible when 'Soil' was selected from the 'Rooting Type' field.
 - Soil Type: drop-down menu - The type of soil: 'Sand', 'Peat', 'Clay', 'Mixed', 'Other'
 - Custom Soil Type: text field - Define a custom soil type. Only visible when 'Other' was selected from the 'Soil Type' field.
 - Soil Container Type: text field - The type of container the soil was in.
- PH Control: drop-down menu - Whether the PH level was controlled during the study: 'controlled', 'uncontrolled'
- PH Value: text field - The PH level during the study. If the level was not controlled, then provide the average PH level.
- Treatments: checkboxes with textfields - The treatments the plants were subject to during the study. If a treatment is checked, a textfield for the description of that treatment will appear below it.

Greenhouse

Greenhouse fields are only visible when 'Greenhouse' was selected from the 'Study Type' field.

- Humidity Control: drop-down menu - Whether the Air Humidity was controlled during the study: 'controlled', 'uncontrolled'
- Humidity Value: text field - The Air Humidity during the study. If the level was not controlled, then the Air Humidity is not required.
- Light Intensity Control: drop-down menu - Whether the Light Intensity was controlled during the study: 'controlled', 'uncontrolled'
- Light Intensity: text field - The Light Intensity during the study. If the level was not controlled, then the Light Intensity is not required.
- Average High Temperature: text field - The average high temperature.
- Average Low Temperature: text field - The average low temperature.
- Rooting Type: drop-down menu - The type of rooting: 'Aeroponics', 'Hydroponics', 'Soil'
- Soil fields are only visible when 'Soil' was selected from the 'Rooting Type' field.
 - Soil Type: drop-down menu - The type of soil: 'Sand', 'Peat', 'Clay', 'Mixed', 'Other'
 - Custom Soil Type: text field - Define a custom soil type. Only visible when 'Other' was selected from the 'Soil Type' field.
 - Soil Container Type: text field - The type of container the soil was in.
- PH Control: drop-down menu - Whether the PH level was controlled during the study: 'controlled', 'uncontrolled'
- PH Value: text field - The PH level during the study. If the level was not controlled, then the PH level is not required.
- Treatments: checkboxes with textfields - The treatments the plants were subject to during the study. If a treatment is checked, a textfield for the description of that treatment will appear below it.

Experimental/Common Garden

Common Garden fields are only visible when 'Common Garden' was selected from the 'Study Type' field.

- Irrigation Type: drop-down menu - The type of irrigation: 'Irrigation from top', 'Irrigation from bottom', 'Drip Irrigation', 'Other', 'No Irrigation'
- Custom Irrigation Type: text field - Define a custom irrigation type. Only visible when 'Other' was selected from the 'Irrigation Type' field.
- Salinity Control: drop-down menu - Whether the Salinity level was controlled during the study: 'controlled', 'uncontrolled'
- Salinity Value: text field - The Salinity level during the study. If the level was not controlled, then provide the average Salinity level.
- Biotic Environment: drop-down menu - The type of biotic environment: 'Herbivores', 'Mutualists', 'Pathogens', 'Endophytes', 'Other', 'None'
- Custom Biotic Environment Type: text field - Define a custom biotic environment type. Only visible when 'Other' was selected from the 'Biotic Environment' field.
- Season: checkboxes - The season(s) that the plants were assessed.
- Treatments: checkboxes with textfields - The treatments the plants were subject to during the study. If a treatment is checked, a textfield for the description of that treatment will appear below it.

Plantation

Plantation fields are only visible when 'Plantation' was selected from the 'Study Type' field.

- Season: checkboxes - The season(s) that the plants were assessed.
- Assessments: drop-down menu - The number of times, on average, the plants were assessed.
- Treatments: checkboxes with textfields - The treatments the plants were subject to during the study. If a treatment is checked, a textfield for the description of that treatment will appear below it.

A screenshot of the Experimental Conditions page can be seen below:

STUDY DESIGN

Data Type: *

Genotype x Phenotype (and/or manual environmental data) x Environmental layers ▼

Study Type: *

Experimental/Common Garden ▼

COMMON GARDEN INFORMATION:

Irrigation Type: *

Irrigation from bottom ▼

Salinity controlled or uncontrolled: *

Uncontrolled ▼

Average Salinity Value: *

Biotic Environment: *

☐ Herbivores

☐ Mutualists

☒ Pathogens

☐ Endophytes

☐ Other

☐ None

Seasons: *

☐ Spring

☒ Summer

☐ Fall

☐ Winter

If you do not know which season your samples were collected, please select all.

4.3 Plant Accession

4.3.1 Study Location

Only common garden studies will have access to these fields.

- Coordinate Projection: drop-down menu - The coordinate projection of the location of the common garden: 'WGS 84', 'NAD 83', 'ETRS 89', 'Custom Location'. If you don't know your coordinate projection, then it is probably WGS 84.

- Study Location(s): Use add/remove location buttons to add/remove locations. Depending on the choice for coordinate projection, this field may only accept properly formatted coordinate locations.

4.3.2 Accession Information

The third set of fields in TPPS is the Plant Accession section. Here you will be asked to submit a file or files that describe each plant with a unique identifier and a geographic location.

- These plants may have been studied in the past: checkbox - If this box is checked, TPPS will try to find plants with matching IDs around the same location as the ones being provided. If it finds them successfully, it will mark them as the same plant in the database.
- Skip location validation (ignore location information): checkbox - Available to administrators only. Checking this box will skip validation of location information for the entire study.
- Plant Accession File: file upload - The plant accession file. This file must have at least a column for Plant ID, and columns describing the location of the plants. There are several options for the location column formats, including GPS coordinates and country/state/district. Files that include more than one species will also need to provide columns for the genus and species of each plant. When the file is uploaded, you will need to define the contents of the file. A table with the header 'Define Data' should appear, where you can select which columns describe the required data. If you do not define the required columns, you will not be able to continue.
- Coordinate Projection: drop-down menu - The coordinate projection of coordinate locations in the plant accession file: 'WGS 84', 'NAD 83', 'ETRS 89'. This feature is currently only available for accession files with GPS coordinate locations.
- After uploading the plant accession file and selecting the correct coordinate projection, you can click the button 'Click here to update map' to view the locations of the plants you described on Google Maps. This can be useful to verify that there are no drastic errors in the locations. This feature is currently only available for accession files with GPS coordinate locations.
- Separate Plant Accession: checkbox - If you would like to upload a separate plant accession file for each species, click this checkbox. This field is only available for studies with more than one species.
- Separate Plant Accession Files: file uploads - The plant accession files. These fields are simply duplicates of the 'Plant Accession File' field above.
- The provided GPS coordinates are exact: checkbox - If the coordinates provided in the Accession file are exact, then check this box. Otherwise, leave it unchecked and fill out the following field.
- Coordinates accuracy: textfield - The precision of the provided coordinates. For example, if a plant could be up to 10m away from the provided coordinates, then the accuracy would be "10m".

A screenshot of the Plant Accession page can be seen below:

Allowed file extensions: txt csv xlsx

tpps_accession_large.xlsx REMOVE

File Upload empty field: NA

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: Plant Identifier and Location. If your plants are located based on a population group, you can provide the population group column and a mapping of population group to location below.

tree identif	Plant Identifier ▼	latitude	Latitude ▼	longitude	Longitude ▼
	TPPS-DELETE-ME-100		42		72
	TPPS-DELETE-ME-101		42		70
	TPPS-DELETE-ME-102		44		71

Please note that if you remain idle on a TPPS page with files for more than 6 hours, you run the risk of having those files reset if you do not click "Save".

☐ My file has no header row

Coordinate Projection WGS 84 ▼



☒ The provided GPS coordinates are exact

4.4 Genotype, Phenotype, and Environment

The fourth set of fields in TPPS is the Genotype, Phenotype, and Environment section. Here you will be asked to provide the data and metadata for your phenotypes and genotypes. If you are submitting a study with multiple species, and would like to use the same phenotype/genotype data across multiple species, you can select the '<Genotype/Phenotype> information for <organism i> is the same as <genotype/phenotype> information for <organism i-1>.' box. Each of the fields on this page is asked once per species:

4.4.1 Phenotype

- Phenotype Metadata: select one or both of the following methods to provide phenotype metadata:
 - Textfields: click the 'Add Phenotype' or 'Remove Phenotype' buttons to add or remove phenotypes, respectively. The following fields are required once per phenotype:
 - * Phenotype Name: text field - The name of the phenotype.
 - * Phenotype Attribute: drop-down menu - The attribute that the phenotype is describing.
 - * Phenotype Custom Attribute: textfield - Custom attribute that was not in the default list.
 - * Phenotype Description: text field - A brief description of the phenotype.
 - * Phenotype Units: text field - The units of the phenotype. For example, "meters", "cm", "inches", "°C", "Degrees Fahrenheit" would be phenotype units.
 - * Phenotype Structure: text field - The structure that the phenotype is describing. Structure can refer to a tissue type or to a biological process. Default is 'Whole plant'.
 - * Phenotype Custom Structure: textfield - Custom structure that was not in the default list.
 - * Phenotype Value Range: text fields - A maximum and minimum value for the phenotype. If your phenotype is binary or has a range, you can click 'Phenotype has a value range' and provide a value range.
 - * Phenotype is an environmental phenotype: checkbox - Check this box to indicate which phenotypes are environmental.
 - File: Click the 'I would like to upload a phenotype metadata file' checkbox to upload a phenotype file. You will be shown a table with several drop-down menus, along with the names of your column headers, and the first few rows of data in your file. You will then be asked to select what type of data each of your columns holds: 'Name/Identifier', 'Attribute', 'Description', 'Units', 'Structure', 'Minimum Value', 'Maximum Value', or 'N/A'. Columns marked 'N/A' will still be kept in the flat file, but will not be recorded in the database with the other data from the file. Columns that hold 'Name/Identifier', 'Attribute', 'Description', 'Units', and 'Structure' must be defined before continuing.
- Phenotype Data: file upload - The phenotype data. This file should contain the 'Plant Id' of the plant that the phenotype is describing, the 'Name/Identifier' of the phenotype, and the 'Value' that was actually measured for this phenotype.

A screenshot of the manual phenotype information fields can be seen below:

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: *

Phenotype 1 Description: *

Please provide a short description of Phenotype 1

Phenotype 1 Units: *

Phenotype 1 Structure: *

☐ Phenotype 1 has a value range

☐ Phenotype 1 is a binary phenotype

☐ Phenotype 1 is an environmental phenotype

A screenshot of the phenotype metadata file field can be seen below:

Phenotype Metadata File: Please upload a file containing columns with the name, attribute, structure, description, and units of each of your phenotypes: *

 tpps_phenotype_meta.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: Phenotype name

Phenotype Name	Phenotype Description	Phenotype Calculation
Phenotype Name/Identifier ▼	Description ▼	Attribute ▼ Units
a2	Per-planted-area cross-sectional area of stems at breast height (1.3m)	Area/area
b2	The number of trees with broken arms	The number of applicable trees in each plot
c2	The number of trees with crooked stems	The number of applicable trees in each plot

Please note that if you remain idle on a TPPS page with files for more than 6 hours, you run the risk of having those files reset if you do not click "Save".

☐ My file has no header row

A screenshot of the phenotype data file field can be seen below:

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

 tpps_phenotype.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: Plant Identifier, Phenotype name, and Value(s)

Site ID	Tree ID	f2	n1	b2
<input type="text" value="N/A"/>	<input type="text" value="Plant Identifier"/>	<input type="text" value="Phenotype Data"/>	<input type="text" value="Phenotype Data"/>	<input type="text" value="Phenotype Data"/>
Site_N	N_6_1_1	0	32	1
Site_N	S_50_1_3	0	15	1
Site_N	N_15_1_4	5	31	2

Please note that if you remain idle on a TPPS page with files for more than 6 hours, you run the risk of having those files reset if you do not click "Save".

☐ My file has no header row

4.4.2 Genotype

- Genotype Marker Type: checkboxes - The types of genotype markers: 'SNPs', 'SSRs/cpSSRs', 'Indels', 'Other'. Select all that apply.
- SNPs Information fields are only visible and required when 'SNPs' was selected from 'Genotype Marker Type'.
 - Experimental Design: drop-down menu - The experimental design: 'GBS', 'Targeted Capture', 'Whole Genome Resequencing', 'RNA-Seq', 'Genotyping Array'
 - GBS Type: drop-down menu - The type of GBS experiment: 'RADSeq', 'ddRAD-Seq', 'NextRAD', 'RAPTURE', 'Other'. Only visible when 'GBS' was selected from the 'Experimental Design' field.
 - Custom GBS Type: text field - Define a custom GBS Type. Only visible when 'Other' was selected from the 'GBS Type' field.
 - Targeted Capture Type: drop-down menu - The type of Targeted Capture experiment: 'Exome Capture', 'Other'. Only visible when 'Targeted Capture' was selected from the 'Experimental Design' field.
 - Custom Targeted Capture Type: text field - Define a custom Targeted Capture type. Only visible when 'Other' was selected from the 'Targeted Capture Type' field.
- SSRs/cpSSRs Type: text field - The type of SSRs/cpSSRs marker. Only visible when 'SSRs/cpSSRs' was selected from the 'Genotype Marker Type' field.
- Other Marker Type: text field - The type of genotype marker. Only visible when 'Other' was selected from the 'Genotype Marker Type' field.

A screenshot of the genotype marker type fields can be seen below:

▼ GENOTYPE INFORMATION:

Marker Type (select all that apply): *

☒ SNPs
☐ SSRs/cpSSRs
☐ Indels
☐ Other

▼ SNPS INFORMATION:

Define Experimental Design: * GBS ▼

GBS Type: * ddRAD-Seq ▼

Reference Assembly used: *

Carica papaya v0.4 ▼

- **Reference Genome/Assembly:** drop-down menu - The reference genome/assembly: TPPS finds local reference genomes from the site it is installed on and their versions, and presents them as options in this drop-down. The other available options for a reference genome/assembly are: 'I can provide a URL to the website of my reference file(s)', 'I can provide a GenBank accession number (BioProject, WGS, TSA) and select assembly file(s) from a list', 'I can upload my own reference genome file', 'I can upload my own reference transcriptome file', 'I am unable to provide a reference assembly'.
- **External Reference Genome/Assembly:** Either upload or locate the genome/assembly file(s).
 - **BioProject Accession number:** text field - The NCBI accession number of the BioProject associated with this study. After this number has been provided, TPPS will search NCBI for assembly files that are linked to this BioProject number and present them to you, as well as links to view them individually on NCBI. You will need to select the assembly files that are relevant to this study. Only visible when 'I can provide a GenBank accession number (BioProject, WGS, TSA) and select assembly file(s) from a list' was selected from the 'Reference Genome/Assembly' field.
 - **Load Reference from URL:** For 'I can provide a URL to the website of my reference file(s)' option, you will be presented with a set of form fields based on the 'Tripal FASTA Loader'. You can skip the 'Existing File(s)' and 'You may select a file that is already uploaded' sections. You will just need to specify the remote path of the reference, the analysis type, and the sequence type. If you would like, you can also specify an external database where the sequences from your url have a cross-reference. If you do, you will also need to provide a regular expression for the accession number. External database cross-reference and accession number are not required.
 - **Upload Reference manually:** For the 'I can upload my own reference genome file', or 'I can upload my own reference transcriptome file' options, you will be presented with a set of form fields based on the 'Tripal FASTA Loader'. You only need to supply a file from the 'Existing File(s)' **or** the 'You may select

a file that is already uploaded' section. You can skip the 'Remote path' section. You will need to specify the assembly type and sequence type. External database for cross-reference and regular expression for external database accession number are optional.

A screenshot of the genotype reference fields can be seen below:

- Ploidy: drop-down menu - Only visible when 'SSRs/cpSSRs Genotype Spreadsheet' is selected from 'Genotype File Types'
- Genotype File Types: checkboxes - 'SNPs Genotype Assay', 'Assay Design', 'SNPs Associations', 'SSRs/cpSSRs Genotype Spreadsheet', 'Indel Genotype Spreadsheet', 'Other Marker Genotype Spreadsheet', 'VCF'. Select all that apply. 'SNPs Genotype Assay', 'SSRs/cpSSRs Genotype Spreadsheet', 'Indel Genotype Spreadsheet', 'Other Marker Genotype Spreadsheet' are only available when 'SNPs', 'SSRs/cpSSRs', 'Indels', 'Other' are selected from the 'Genotype Marker Type' field, respectively. 'Assay Design' and 'SNPs Associations' are only available if 'SNPs Genotype Assay' is selected.
- SNPs Genotype Assay File: file upload - The SNPs genotype assay file. The format of this file is very important! The first column should contain plant identifiers which match the plant identifiers provided in the accession file and all of the remaining columns should contain SNP data.
- Assay Design File: file upload - design file for SNPs genotype assay.
- SNPs Associations File: file upload - Spreadsheet containing SNPs Association data. You will be required to specify which of your columns holds the SNP ID, scaffold, position, allele, associated trait, and confidence value.
 - Confidence Value Type: drop-down menu - The type of confidence value, such as P-value, Genomic Inflation Factor, etc.

- Association Analysis Tool: drop-down menu - Select analysis tool from list of options.
- SNPs Population Structure file: optional file upload.
- SNPs Kinship File: optional file upload.
- SSRs/cpSSRs Spreadsheet: file upload - SSRs/cpSSRs genotype file. The format of this file is very important! The first column should contain plant identifiers and each remaining column will be SSR/cpSSR data. The exact format depends on the ploidy selected.
 - Extra SSRs/cpSSRs Spreadsheet: checkbox - Check this box to add an additional SSRs/cpSSRs file.
 - Additional SSRs/cpSSRs Type: textfield - Marker type of the additional SSRs/cpSSRs marker.
 - Additional SSRs/cpSSRs Ploidy: Ploidy of the additional SSRs/cpSSRs marker.
 - SSRs/cpSSRs Additional Spreadsheet: file upload - Additional SSRs/cpSSRs spreadsheet. Same format as the first SSRs/cpSSRs spreadsheet.
- Indel Genotype Spreadsheet: file upload - Indel genotype spreadsheet. The format of this file is very important! The first column should contain plant identifiers and all remaining columns should contain indel data.
- Other Marker Genotype Spreadsheet: file upload - Other marker genotype spreadsheet.
- VCF File: file upload - The VCF file. Must be a .zip file.

A screenshot of the genotype file fields can be seen below:

Genotype File Types (select all that apply): *

☒ SNPs Genotype Assay
☐ Assay Design
☐ SNPs Associations
☒ VCF

SNPs Genotype Assay File: please provide a spreadsheet with columns for the Plant ID of genotypes used in this study: *

No file chosen

Please upload a spreadsheet file containing SNP Genotype Assay data. The format of this file is very important! The first column of your file should contain plant identifiers which match the plant identifiers you provided in your plant accession file, and all of the remaining columns should contain SNP data.

Genotype VCF File: *

No file chosen

4.4.3 Environment

- Use CartograPlant Layers: If CartograPlant Layers were enabled by your site administrator, you should have the option to use Environmental Layers provided by CartograPlant that you used in your study.
- CartograPlant Environmental Layers: checkboxes - All available CartograPlant Environmental Layers. From here you can view the names and descriptions of each CartograPlant Environmental layer you have to choose from.

- CartograPlant Environmental Layer Parameters: checkboxes - If you selected some CartograPlant Environmental Layers, you will need to indicate which variables from the layer you used in your study. Parameter fields will only be visible for layers that you selected.
- Custom Environmental Layer: checkbox - Non-CartograPlant environmental layer. Specify the database URL, the name of the layer, and the parameters used.

4.5 Submission Summary

Once all of the data on the previous pages has been submitted, a brief overview of the data provided so far is shown. A sample screenshot of the summary can be seen below.

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

Publication Title	TPPS Example Submission
Primary Author	Peter Richter
Organization	University of Connecticut
Publication Status	In Press
Publication Abstract	Abstract placeholder for Example TPPS Submission.
Publication Journal	Journal of bioscience and bioengineering
Species 1	Pinus taeda
Experiment Starting Date	July 1973
Experiment Ending Date	June 1984
Study Location	41.8077° N, 72.2540° W
Data Type	Genotype x Phenotype x Environment
Study Type	Natural Population (Landscape)
Seasons Assessed	Spring Summer
Average number of assessments	15
Tree Accession File	ttps_accession_test.xlsx
Pinus taeda Phenotype Metadata File	phenotype_metadata.xlsx
Pinus taeda Phenotype File	phenotype_file.xlsx
Pinus taeda Genotype Marker Types	SNPs
Pinus taeda Genotyping Design	GBS
Pinus taeda GBS Type	ddRAD-Seq
Pinus taeda Uploaded Reference File	661
Pinus taeda Genotype VCF File	test.vcf
Cartogratree Layer: Minimum temperature January Parameters	Minimum temperature
Cartogratree Layer: Minimum temperature February Parameters	Minimum temperature

Optional fields for comments and supplemental files are available on the summary page. A screenshot of those fields can be seen below

If you have any additional comments about this submission you would like to include, please write them here:

▼ ADDITIONAL FILES

If there are any additional files you would like to include with your submission, please upload up to 10 files here.

ADD FILE

REMOVE FILE

Supplemental File 1

Choose File No file chosen

UPLOAD

Supplemental File 2

Choose File No file chosen

UPLOAD

5.1 Coding Style Specifications

TPPS usually tries to adhere to the *Drupal coding style standards*. We will cover some of the most important coding style standards here:

- Tabs are 2 spaces, indent using spaces not tabs.
- Lines should never end in white space.
- Starting curly braces should be on the same line as the if/for/while/etc statement, preceded by a space.
- Ending curly braces should have their own line.
- Each .php and .inc file should have a comment block describing the file. See any TPPS .php or .inc file for an example.
- Each function in a .php or .inc file should have a comment block describing the function, as well as all of the parameters, return values, and global variables that the function uses.
- Use Tripal API functions instead of Drupal API functions.
- Variable names should use snake_case.
- Function names in php should use snake_case, in JavaScript should use camelCase.

5.2 Project Structure

```
- admin
  - config.php
  - panel.php
- ajax
  - tpps_ajax.php
- css
  - tpps.css
```

(continues on next page)

(continued from previous page)

```
- forms
  - build
    - front.php
    - page_1.php
    - page_1_ajax.php
    - page_1_helper.php
    - page_2.php
    - page_2_ajax.php
    - page_2_helper.php
    - page_3.php
    - page_3_ajax.php
    - page_3_helper.php
    - page_4.php
    - page_4_ajax.php
    - page_4_helper.php
    - summary.php
  - validate
    - page_1.php
    - page_2.php
    - page_3.php
    - page_4.php
  - submit
    - submit_all.php
- includes
  - accession_coordinates.inc
  - compare_files.inc
  - completed_display.inc
  - create_record.inc
  - cron.inc
  - file_parsing.inc
  - flatten.inc
  - get_env_data.inc
  - init_project.inc
  - manage_doi.inc
  - parse_xlsx.inc
  - save_file_columns.inc
  - standard_coord.inc
  - status_bar.inc
  - submissions.inc
  - submit_email.inc
  - tab_create.inc
  - table_display.inc
  - validate_columns.inc
  - zenodo.inc
- js
  - tpps.js
- tests
  - bootstrap.php
  - DataFactory.php
  - example.env
  - ProjectInitTest.php
- tpps.info
- tpps.module
- tpps.install
- README.md
- LICENSE
- composer.json
```

(continues on next page)

(continued from previous page)

```
- composer.lock
- phpunit.xml
```

The admin/ folder contains code that build the forms to manage TPPS settings and TPPS submissions. The ajax/ folder contains code for ajax callback functions that need to be accessible to any part of the TPPS module. The css/ folder contains stylesheets for TPPS, and the js/ folder contains JavaScript that needs to be accessible to any part of the TPPS module. The forms/ folder contains code for functions that build, validate, and submit the main TPPS form. The includes/ folder contains code for helper functions that are used throughout the TPPS module. The tests/ module contains code for unit tests that are run automatically by TravisCI every time the code is pushed. The TPPS project on TravisCI can be found at [this link](#).

5.3 File Manipulation

File Manipulation in TPPS is a complex topic, because TPPS needs to provide file uploads that can be parsed easily, but also need to be very flexible. TPPS achieves this by allowing users to define the data types of each of the columns in a file upload. For example, for the plant accession file, users need to define columns that describe the plant identifier and location of that plant identifier. Creating the file upload field as well as the fields to define column types and show a preview of the data inside the file requires a lot of code, so some parts of it are found in the normal form build files, some parts are found in hooks in the tpps.module file, and some parts of it are found in helper functions in the tpps.module file. Here we will attempt to cover all aspects of the file manipulation process.

For reference, we will be using the Plant Accession file field as our example through the entire process.

5.3.1 Define Drupal Form Element

The first part of creating a dynamic TPPS file upload field is specifying it in the \$form array just as you would for any other Drupal Form API field:

```
$form['tree-accession']['file'] = array(
  '#type' => 'managed_file',
  '#title' => t("Plant Accession File: *<br>$file_description"),
  '#upload_location' => "$file_upload_location",
  '#upload_validators' => array(
    'file_validate_extensions' => array('txt csv xlsx'),
  ),
  '#states' => ($species_number > 1) ? (array(
    'visible' => array(
      ':input[name="tree-accession[check]"'] => array('checked' => FALSE),
    ),
  )) : NULL,
  '#field_prefix' => '<span style="width: 100%;display: block;text-align: right;
padding-right: 2%;">Allowed file extensions: txt csv xlsx</span>',
  '#suffix' => '<style>figure {}</style>',
);
```

The type of Drupal Form API field we will be building on is the `managed_file` field. This is the closest contributed field type to what TPPS needs. If the field definition is left like this, then it will operate the same as any other managed file field. This is important because there are a few file uploads within the TPPS form that do not require the user to define their column data, so we want to preserve the same managed file functionality for those fields.

Now we will add attributes to the file field which will include some additional information about how this specific dynamic file upload field should behave:

```
$form['tree-accession']['file']['empty'] = array(
  '#default_value' => isset($values['tree-accession']['file']['empty']) ? $values[
  ↳ 'tree-accession']['file']['empty'] : 'NA',
);

$form['tree-accession']['file']['columns'] = array(
  '#description' => 'Please define which columns hold the required data: Plant_
  ↳ Identifier and Location. If your plants are located based on a population group,
  ↳ you can provide the population group column and a mapping of population group to
  ↳ location below.',
);

$column_options = array(
  '0' => 'N/A',
  '1' => 'Plant Identifier',
  '2' => 'Country',
  '3' => 'State',
  '4' => 'Latitude',
  '5' => 'Longitude',
  '8' => 'County',
  '9' => 'District',
  '12' => 'Population Group',
);

...

$form['tree-accession']['file']['columns-options'] = array(
  '#type' => 'hidden',
  '#value' => $column_options,
);

$form['tree-accession']['file']['no-header'] = array();
```

Now, in addition to our plain managed file field, we have defined some subfields: an incomplete field named `empty` with a default value, an incomplete field named `columns` with a description, a hidden field named `columns-options` with an associative array of the data types we would like the user to choose from when defining their data, and an empty field named `no-header`. These fields are important to the next section, where some hook functions and helper functions will interpret them and use them to create the column dropdowns.

5.3.2 TPPS Element Info Alter

The hook function named `tpps_element_info_alter` makes changes to form elements of certain types. If you would like to learn more about the hook `hook_element_info_alter()`, you can visit [this link](#). Right now we are only interested in the changes it makes to `managed_file` element types:

```
function tpps_element_info_alter(&$type) {
  ...

  $type['managed_file']['#process'][] = 'tpps_managed_file_process';
}
```

The function appends the name of a new callback function to the end of the managed file `#process` attribute. This means that when the form is processed, the function named `tpps_managed_file_process` will be called after each of the managed file's other process functions.

5.3.3 TPPS Managed File Process

The `ttps_managed_file_process` function is pretty complicated, so we'll be breaking it down into smaller parts and explaining them here. The first snippet ensures that no changes are made to the managed file field if it is not inside the main TPPS form:

```
function ttps_managed_file_process(array $element, array &$form_state, array $form) {
  if ($form_state['build_info']['form_id'] !== 'ttps_main') {
    return $element;
  }
}
```

The next section checks whether the `columns` attribute of the file has been defined, then loads some information that will be needed later, and defines the no-header checkbox. The important variables that are defined here are `$wrapper` and `$fid`. `$wrapper` is the id of the `div` element that wraps the file field. This is important for ajax functions. `$fid` is the id of the managed file that the user has uploaded. If the user has not uploaded a file yet, then `$fid` will be 0:

```
if (isset($element['columns'])) {
  require_once 'ajax/ttps_ajax.php';
  $fid = $element['#value']['fid'];
  $wrapper = substr($element['#id'], 0, -7) . '-ajax-wrapper';

  $saved_value_parents = $no_header_parents = $element['#parents'];
  $no_header_parents[] = '#value';
  $no_header_parents[] = 'no-header';

  $no_header = drupal_array_get_nested_value($form_state['complete form'], $no_header_
↪parents);
  $callback = isset($form_state['triggering_element']['#ajax']['callback']) ? $form_
↪state['triggering_element']['#ajax']['callback'] : NULL;
  if (!$no_header and ($callback != 'ttps_no_header_callback')) {
    $end = array_pop($saved_value_parents);
    $saved_value_parents[] = $end . "-no-header";
    $no_header = drupal_array_get_nested_value($form_state['saved_values'][$form_
↪state['stage']], $saved_value_parents);
  }

  $element['no-header'] = array(
    '#type' => 'checkbox',
    '#title' => 'My file has no header row',
    '#ajax' => array(
      'wrapper' => $wrapper,
      'callback' => 'ttps_no_header_callback',
    ),
    '#states' => isset($element['#states']) ? $element['#states'] : NULL,
    '#default_value' => $no_header ? $no_header : NULL,
  );
};
```

The next snippet will check if the user has selected a file. If they have, then TPPS attempts to load it. If it loads correctly, then TPPS starts to define the `columns` fieldset which will hold each of the column data type fields. TPPS will then stop usage of the file (so that it may be deleted if the user clicks the remove button) and load the file uri, location of the file, the content of the file, and the column data type options:

```
if (!empty($fid) and ($file = file_load($fid))) {

  $saved_vals = $form_state['saved_values'][$form_state['stage']];
  $element['columns']['#type'] = 'fieldset';
```

(continues on next page)

(continued from previous page)

```
$element['columns']['#title'] = t('<div class="fieldset-title" style="font-size:.8em
→">Define Data</div>');
$element['columns']['#collapsible'] = TRUE;

$file_name = $file->uri;

// Stop using the file so it can be deleted if the user clicks 'remove'.
file_usage_delete($file, 'tpps', 'tpps_project', substr($form_state['accession'],
→4));

$location = drupal_realpath("$file_name");
$content = tpps_parse_xlsx($location, 3, !empty($no_header));

$options = $element['columns-options']['#value'];
```

Now TPPS will define the actual column data type drop down fields:

```
$first = TRUE;

foreach ($content['headers'] as $key => $item) {
  $item_parents = $element['#parents'];
  array_pop($item_parents);
  $item_parents[] = end($element['#parents']) . '-columns';
  $item_parents[] = $key;

  $default = drupal_array_get_nested_value($saved_vals, $item_parents);
  // dpm($item_parents);
  $element['columns'][$key] = array(
    '#type' => 'select',
    '#title' => $item,
    '#options' => $options,
    '#default_value' => $default,
    '#prefix' => "<th>",
    '#suffix' => "</th>",
    '#attributes' => array(
      'data-toggle' => array('tooltip'),
      'data-placement' => array('top'),
      'title' => array("Select the type of data the '$item' column holds"),
    ),
  );

  if ($first) {
    $first = FALSE;
    $first_col = $key;
  }

  if (!empty($no_header)) {
    $element['columns'][$key]['#title'] = '';
    $element['columns'][$key]['#attributes']['title'] = array("Select the type of
→data column $item holds");
  }
}
```

Finally, TPPS will display a preview of the data found in the file and populate the subfield named `empty` if it exists:

```
$rows = $content;
unset($rows['headers']);
```

(continues on next page)

(continued from previous page)

```

$headers = array();
foreach ($content['headers'] as $col_name) {
    $headers[] = $col_name;
}
$vars = array(
    'header' => $headers,
    'rows' => $rows,
    'attributes' => array('class' => array('view')),
    'caption' => '',
    'colgroups' => NULL,
    'sticky' => FALSE,
    'empty' => ''
);
$table = theme_table($vars);
preg_match('/\A(.*)<thead[A-Z|a-z|"|\'|_|0-9]*>.*(</thead>.*</table>)/s',
↪$table, $matches);

$element['columns'][$first_col]['#prefix'] = "<div style=\"overflow-x:auto\">" .
↪ $matches[1] . "<tr>" . $element['columns'][$first_col]['#prefix'];
$element['columns'][$key]['#suffix'] = "</tr>" . $matches[2] . "</div>";
}
}

if (isset($element['empty'])) {
    $element['empty']['#type'] = 'textfield';
    $element['empty']['#title'] = t('File Upload empty field:');
    $element['empty']['#states'] = isset($element['#states']) ? $element['#states'] :
↪NULL;
    $element['empty']['#description'] = 'By default, TPPS will treat cells with the
↪value "NA" as empty. If you used a different empty value indicator, please provide
↪it here.';
}

return $element;
}

```

The returned element should now be a fully populated TPPS dynamic file upload field.

5.3.4 Values and Parsing

We will now discuss the format of TPPS Dynamic file values and the ways in which they are parsed and validated.

5.4 Where Does Data Go?

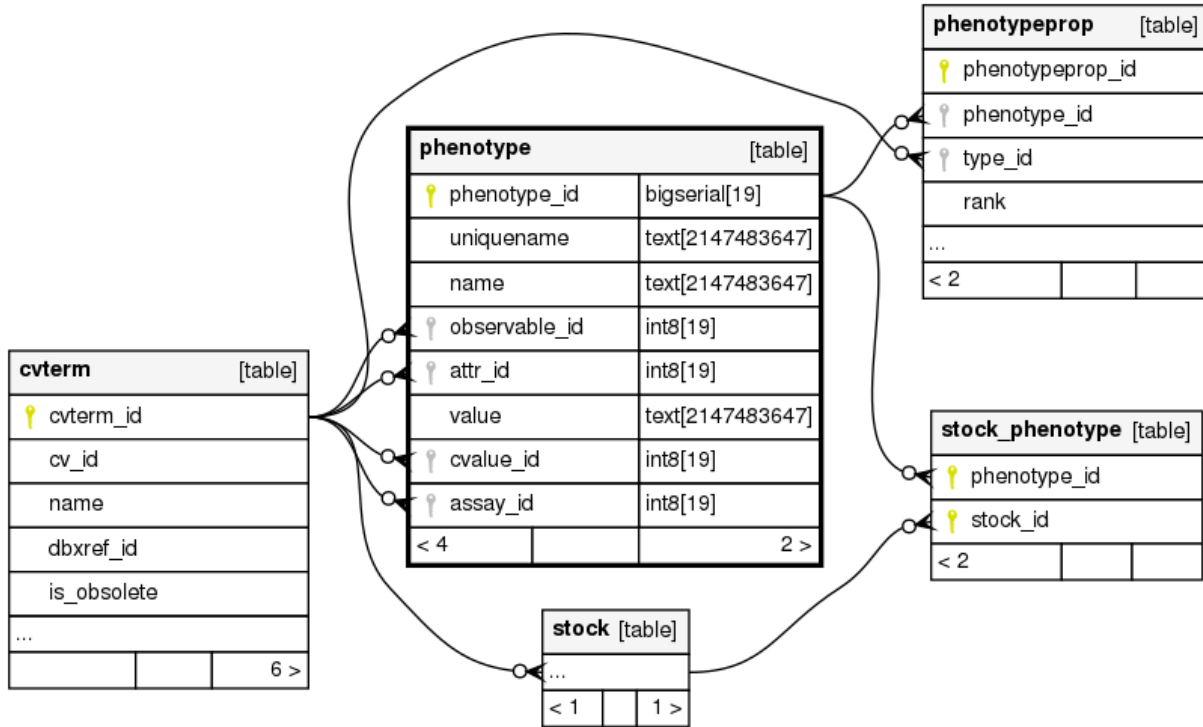
This section assumes that the reader is familiar with the [Chado](#) relational database schema. It will explain where data is stored and how different data types within the same TPPS submission are associated with one another.

5.4.1 Phenotype

For each phenotype measurement, a phenotype record is created with a unique name with the format <Project Accession>-<Plant ID>-<Phenotype Name>. The value is the measured value of the phenotype (or

null if value is not applicable). The metadata for the phenotype (unit, description, min/max values, etc) is then represented with associated `phenotypeprop` records. The phenotype is associated with the plant through the `stock_id` of the base `phenotype` record, as well as through the `stock_phenotype` table.

Here is a representation of the relationships between some of the phenotype tables for reference:



Generated by SchemaSpy

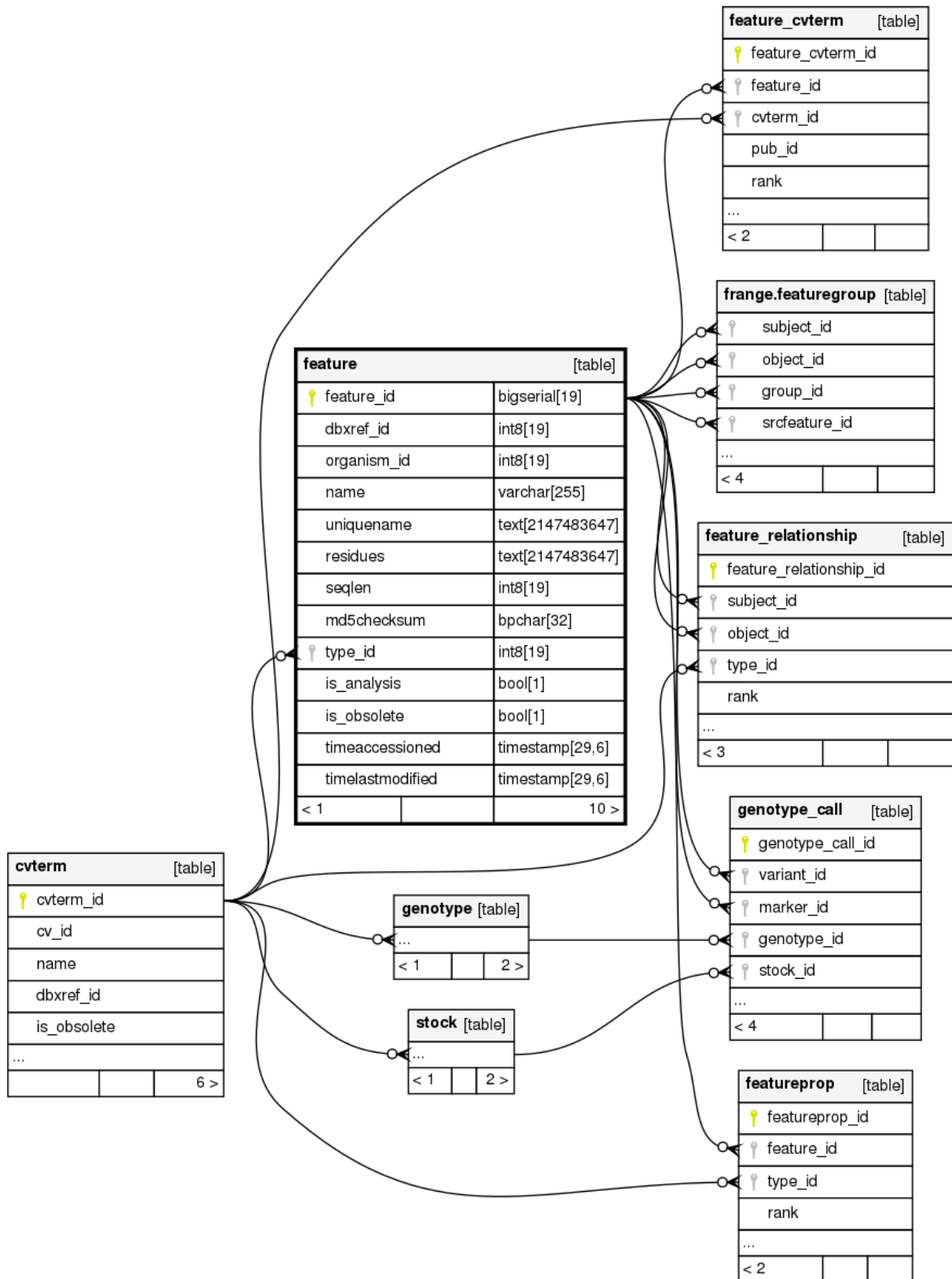
5.4.2 Genotype

How genotype data gets submitted in TPPS varies slightly based on the type of genotype data and the format it is provided in, but the locations in Chado and the associations between the data stay almost entirely the same. The function that records all of this data can be found [here](#).

For each genotype marker, two separate `feature` records are created, called the `marker` and `variant` records. In addition to the `feature` records, a `genotype` record is created. Each genotype call for each individual is represented in the `genotype_call` table, which associates the marker, variant, genotype, stock, and project together and also holds any additional metadata about the call. Each `stock` is also associated with each genotype in the `stock_genotype` table.

If association data is available for a genotype marker, then an additional `feature` record is created with a `feature_cvterm` for the `cvterm` of the associated trait and a `featureprop` for the type of association statistics being used (P-value, Bonferroni, etc). A `feature_relationship` record will then be created between the association feature and the `variant` feature, and the value of this record will be the calculated association confidence value.

Here is a representation of the relationships between some of the genotype tables for reference:



Generated by SchemaSpy