
GTTN-DR Documentation

Peter Richter

May 05, 2020

Contents:

1	Introduction	1
2	Overview	3
2.1	Genotype and Phenotype Data	4
2.2	Georeferenced Accessions	4
2.3	Method-Specific Metadata	4
2.4	Data Access Options	4
3	Features	5
3.1	Data Types and Standards	5
3.2	Data Accessibility	5
3.3	User Friendliness	5
3.4	Administrative Features	6
4	User Information	7
4.1	Data Collection Pipeline	7
4.2	Reference and Sample Data Search Form	22
5	Administrator Information	27

CHAPTER 1

Introduction

A collaboration between University of Connecticut's Plant Computational Genomics lab and the Global Timber Tracking Network. Allows for secure and convenient submission of multiple tree descriptors through a web-based interface.

This module is a modified extension of the Tripal Plant PopGen submit Pipeline (TPPS), which can be found here: <http://tpps.rtd.io>

This module is currently in demo form, meaning that the data is not submitted to the TreeGenes database, and there is an additional "results" page, which displays the data in short text when the user clicks "Submit".

The first page of the module prompts the user for information about each species they are uploading data about, as well as a file with location information and unique identifiers for each tree.

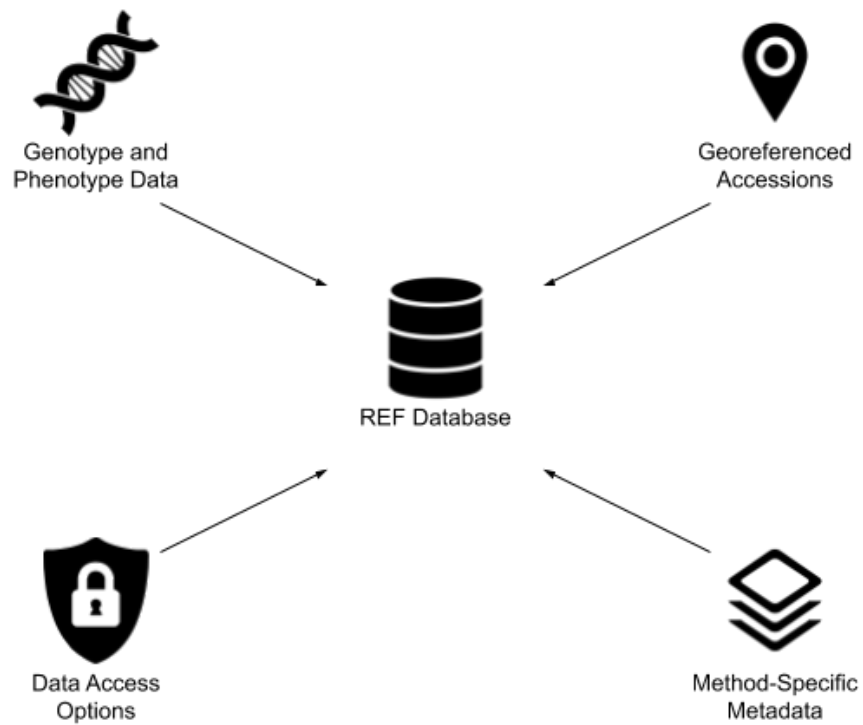
The second page of the module asks for information about the sampling and analysis dates of each species, as well as any phenotypic or genotypic data files the users have.

Unlike TPPS, this form can only be accessed by members of the 'gttn' or 'administrator' groups on the TreeGenes site.

CHAPTER 2

Overview

GTTN-TPPS is a data collection tool developed with the goal of collecting high-quality reference data for the purposes of timber tracking and identification. The module collects 4 different types of data:



2.1 Genotype and Phenotype Data

The core of the data that GTTN-TPPS will collect is the Genotype and Phenotype data. The genotype data might include data in the form of SNPs, genotyping assays, SSRs, etc. The phenotype data might include data in the form of DART data, wood anatomy, etc. DART data could include a series of files with peak data for various isotopes found in a DART scan. Wood anatomy data could include images of microscope slides, along with the specific anatomical features found within the slide in question. Several organizations have expressed interest in providing reference data to GTTN, including:

- [Royal Botanic Gardens, KEW](#) (Wood Anatomy)
- [Ghent University/Royal Museum for Central Africa, Belgium](#) (DART Isotope)
- [Agroisolab UK](#) (Stable Isotope)
- [Thünen Institute of Forest Genetics](#) (Genetic Markers)

2.2 Georeferenced Accessions

When the Genotype and Phenotype Data are submitted through GTTN-TPPS, it will come along with georeferenced accessions of the trees which were sampled in order to obtain the data. Georeferenced Accessions will usually be submitted in the form of an excel table, mapping each tree identifier to a latitude/longitude coordinate. The georeferenced accessions will be integrated with the Genotype and Phenotype Data in the REF Database.

2.3 Method-Specific Metadata

In addition to Georeferenced Genotype and Phenotype Data, each GTTN-TPPS submission will also include metadata which is specific to the analysis method used to obtain the data. For example, if the submission includes DART data, then part of the metadata GTTN-TPPS will collect might be the settings of the DART machine used to obtain the data, or if the submission includes Genotyping by Sequencing data, then part of the metadata GTTN-TPPS will collect might include the type of Genotyping by Sequencing: ddRAD, RAD, NextRAD, etc. For more details, you can view this [Metadata Document](#) which was put together at the March 2019 GTTN workshop in Koli, Finland.

2.4 Data Access Options

The 4th type of data that will be collected in a GTTN-TPPS submission is the Data Access and Authorization Options. Here users will be allowed to select which organizations within the GTTN network are allowed to see the data being submitted, whether the data will be published to TreeGenes, etc.

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

3.1 Data Types and Standards

- Support for genotype and phenotype data and metadata
- 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

3.2 Data Accessibility

- Data is standardized and stored in the local database so that other tools, for example, [CartograTree](#), can easily collect and analyze it
- Restricted access to users with the specific *gttn* user group.
- The studies can be queried or downloaded (flatfiles) through the Tripal interface
- Display both complete and incomplete submissions on ‘GTTN-TPPS Submissions’ user profile tab

3.3 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

3.4 Administrative Features

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

4.1 Data Collection Pipeline

This section contains details on how to get access and use the GTTN-TPPS Data Collection Pipeline.

4.1.1 Creating an Account

Before we can start submitting data through the form, we must create an account so that GTTN-TPPS knows who is submitting data and which organization that data is coming from. To create an account, navigate to gttn.treegenesdb.org/user/register.

You will be asked to provide your full name, your email address, and indicate which organizations you are a part of:

Create new account

Log in

Request new password

E-mail address *

A valid e-mail address. All e-mails from the system will be sent to this address. The e-mail address is not made public and will only be used if you wish to receive a new password or wish to receive certain news or notifications by e-mail.

Full name:

Please indicate the organizations you are part of:

☐ University of Connecticut

☐ EFI

CREATE NEW ACCOUNT

After you have provided all of the required information, an admin will need to approve your account and the primary contacts of each organization you indicated will need to verify that you are part of that organization. Once your account has been approved and verified, you should receive an email notification and you will be able to set a password and log in.

Once you are logged in, you will have access to a variety of new data, depending on which organizations you claimed membership. Data that is public to GTTN organizations, data that has been shared with organizations that you are a member of, and data that has been shared with your user roles will now become available.

4.1.2 User Roles

There are three default user roles that are important to understanding data access, and there are also some special roles that give users special permissions. We will discuss all of the roles currently available on gtn.treegenesdb.org here:

Default user roles:

- **Anonymous:** This role is automatically assigned to anyone who is not yet logged in to the GTTN site. This will restrict data access to collections that have been marked as available to the public.
- **Authenticated:** This role is assigned to anyone who is now logged in to the GTTN site. Having this role allows the user to see all of the same information as the “Anonymous” role, as well as collections that have been marked as available to the entire GTTN network, **and** collections that have been marked as available to organizations which the user is a part of.
- **Administrator:** This role is only given to site administrators, usually limited to those who work on developing the code of the site. This role allows users to browse all data on the site, regardless of membership in an organization.

Custom user roles:

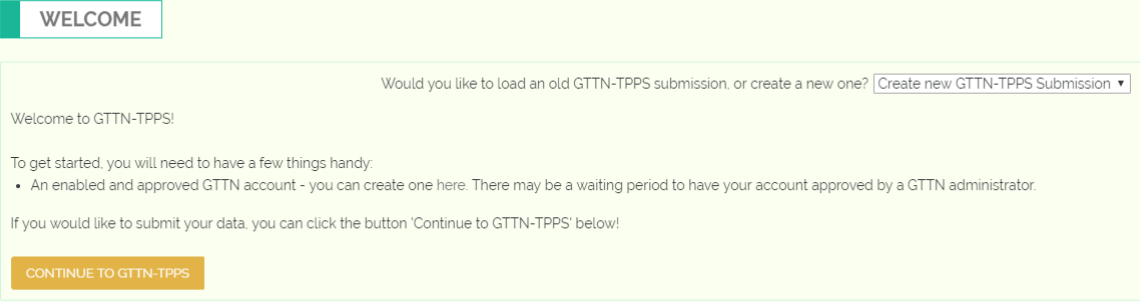
These roles are for users who serve a special purpose in the GTTN community, and usually result in access to additional data:

- **Law Enforcement:** This role describes users who are verified law enforcement officers. Having this role allows the user to see all of the same information as the “Authenticated” role, as well as collections that have been marked as available only to law enforcement users.

Please note that the access that these roles provide is subject to change, and additional custom user roles are likely to be added in the future!

4.1.3 Landing Page and Submission Type

Now that you have created an account and have successfully logged in, you can start submitting data through the data collection pipeline! Navigate to gttn.treegenesdb.org/gttn-tpps in your browser and you will see the landing page. If this is your first time submitting data then your only option will be to create a new GTTN-TPPS Submission, but if you have saved incomplete submissions, you will be able to choose to load one of those submissions:



Click “Continue to GTTN-TPPS” and you will see the Submission Type page. This page collects some metadata about the data submission and the higher level project funding, where applicable:

✖ Species Information
✖ Tree Accession
✖ Submit Data

PROJECT BASIC INFORMATION

Submission Name: *

Data Collection Purpose: *

Please provide a brief description of why this data was collected.

PROJECT BACKGROUND

Which of your organization are you submitting for?

☐ test-organization-1
☐ test-organization-2

Analysis Date: *

Apr

20

2020

Publication DOI:

Data DOI:

Original Database URL:

Project Name (Funding Agency/Grant Number):

Submission Type: *

- Select -

Data permissions

☒ test-organization-1
☒ test-organization-2

Please select the organizations which are allowed to view or browse this data

☐ I have read and agree to the following disclaimer: *
This is the placeholder disclaimer

NEXT

Required fields will be marked with an asterisk. The Submission name under Project Basic Information is the name of the data collection that you will be submitting, for example if you were submitting a DART analysis, you might name your submission something like “DART Analysis 1”. The Project Name under the Project Background section is the name of the higher level project or the NSF grant number. This field is not necessary but can make your data easier to find in the future.

If you are a member of more than one organization, you will need to indicate which organization you are submitting this data for. This is important for keeping track of the sample inventory of each organization in the GTTN group.

GTNN needs to know the type of trees and samples you will be submitting - if they already exist in the GTTN database or if they are brand new. This information is collected through the “Submission Type” field.

You will then need to select which organizations are allowed to see the data in the “Data Permissions” field. If you opt not to select any organizations, then the data you provide through this submission will be visible only to you.

4.1.4 Species and Data Type Information

This page will allow you to specify the species associated with your data and the data types you will be submitting:

⚠ Species Information
✖ Tree Accession
✖ Submit Data

ORGANISM INFORMATION:

ADD ORGANISM
REMOVE ORGANISM

Species 1

Please select your species from the autocomplete list. If your species is not in the autocomplete list, then a new species will be added to the database.

DATA TYPES

Data Type: *

☐ Sample Data
☐ DART Reference Data
☐ Isotope Reference Data
☐ Genetic Reference Data
☐ Anatomical Reference Data

*: Required Field

BACK
SAVE
NEXT

You can add as many species to your submission as you would like. The species fields will autocomplete your entries to species that are already present in the database. If your species is not in the database, that's ok! When you complete your submission a new species will be added.

You will then need to select all of the data types you will be submitting. Your options are Sample, DART, Isotope, Genetic, and Anatomy data. Your selection here will influence the fields you will see in later in the form.

4.1.5 Location Information

This page will collect location information for the trees you are using in this submission, as well as sample information if you indicated you were providing sample data in the previous page.

✓ Species Information ⚠ Tree Accession ✗ Submit Data

TREE ACCESSION INFORMATION

Entandrophragma cylindricum Accession File: *

Please upload a spreadsheet file containing tree population data. When your file is uploaded, you will be shown a table with your column header names, several drop-downs, and the first few rows of your file. You will be asked to define the data type for each column, using the drop-downs provided to you. If a column data type does not fit any of the options in the drop-down menu, you may omit that drop-down menu. Your file must contain columns with information about at least the Tree Identifier and the Location of the tree (either gps coordinates or country/state).

Allowed file extensions: txt csv xlsx

Choose File

No file chosen

UPLOAD

File Upload empty field: NA

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

☐ My file has no header row

Coordinate Projection: WGS 84

[CLICK HERE TO VIEW TREES ON MAP!](#)

SAMPLE INFORMATION

☒ These samples are physical samples

Sample File: *

Allowed file extensions: txt csv xlsx

Choose File

No file chosen

UPLOAD

File Upload empty field: NA

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

☐ My file has no header row

Sample Collection Date: *

May 5 2020

Sample Collector: *

The person who collected the samples.

Tissue Type: - Select -

Sampling Method: - Select -

☐ These samples have been analyzed

Storage location: University of Connecticut

☐ These samples can be shared

Tree Accession Information

This section will require one or more tree accession files:

TREE ACCESSION INFORMATION

Entandrophragma cylindricum Accession File: *

Please upload a spreadsheet file containing tree population data. When your file is uploaded, you will be shown a table with your column header names, several drop-downs, and the first few rows of your file. You will be asked to define the data type for each column, using the drop-downs provided to you. If a column data type does not fit any of the options in the drop-down menu, you may omit that drop-down menu. Your file must contain columns with information about at least the Tree Identifier and the Location of the tree (either gps coordinates or country/state).

Allowed file extensions: txt csv xlsx

accession_small.xlsx

REMOVE

File Upload empty field: NA

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

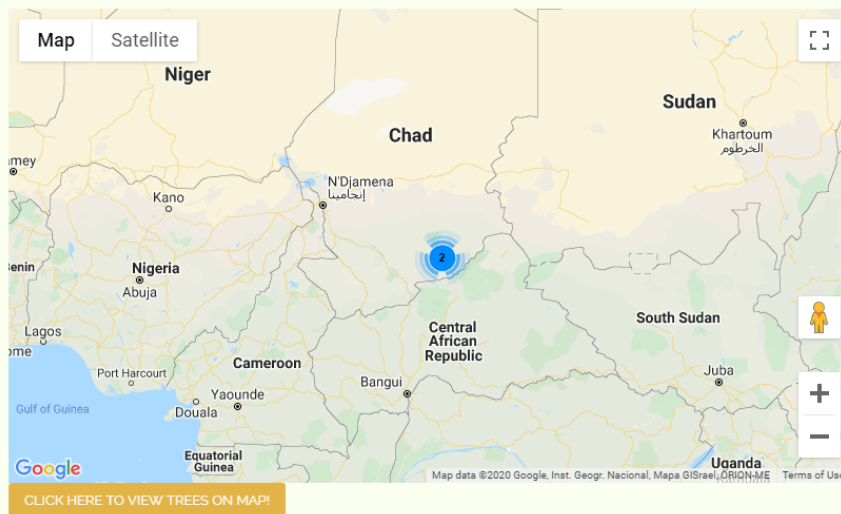
DEFINE DATA

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

tree identifier	Tree Identifier	latitude	Latitude	longitude	Longitude
	TPPS-DELETE-ME-100		10		20
	TPPS-DELETE-ME-101		11		19

☐ My file has no header row

Coordinate Projection | WGS 84



The simplest tree accession file will simply require a tree identifier column, and location columns such as latitude/longitude, or country/state. However, if your accession file contains trees from multiple species, you will also need columns that indicate which species a tree is from. If you do not have the exact locations for your trees and instead have population groups, you can indicate a population group column, and you will be prompted to indicate the location of each population group below the file field.

If you have properly filled out the column types for tree identifier and location, you will be able to view your data in a thumbnail map by clicking “Click here to view trees on map!”. This is useful for verifying that the locations you have uploaded are being interpreted correctly by GTTN-TPPS.

Sample Information

This section requires a sample file:

SAMPLE INFORMATION

☒ These samples are physical samples

Sample File: *

Allowed file extensions: txt csv xlsx

gttn_samples.xlsx

REMOVE

File Upload empty field: NA

By default, GTTN-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

internal sample id	tissue type	remaining volume of sample	sample collector	sample source
Internal Sample ID	Tissue Type	Remaining Volume of Sample	Sample Collector	Sample Source
Tw58	bark	10cm ³	Peter Richter	TF
Tw207	bark	20cm ³	Peter Richter	
Tw218	bark	30cm ³	Peter Richter	TI

☐ My file has no header row

Sample Collection Date: *

May

5

2020

☒ These samples have been analyzed

Storage location: University of Connecticut

☒ These samples can be shared

This section requires a variety of information about each sample you are submitting. The sample file you provide must contain the following columns:

- Internal Sample ID or Xylarium ID
- Sample Source - Which tree or other sample does this sample come from?
- Sample Dimensions - What are the LxWxH dimensions of the sample?
- Remaining Volume of Sample - How much of the sample is left?

You will then need to provide this information in either a file column or by filling out the fields below the file upload field:

- Collection Date - The date the sample was collected
- Sample Collector - The person who collected the sample
- Sample Tissue - The type of tissue the sample is. This is usually bark, heartwood, leaf, etc.
- Sampling Method - The method of collecting the sample. This is either increment core, punch, disc, or cube.
- Analyzed - Whether or not the sample has already been analyzed.
- Storage Location - The location where the sample is being stored. This is how GTTN keeps track of the inventory of each organization and the location of each sample.

Finally, you will need to indicate whether the samples can be shared with other organizations.

4.1.6 Reference and Analysis Data

This page contains reference and analysis data. The contents of the form depend on the selections made on the Species and Data Type Information page.

Direct Analysis in Real Time (DART)

This section collects information about a DART analysis:

DART REFERENCE DATA

DART Reference Metadata File: *

Allowed file extensions: txt csv xlsx
DART_test.xlsx REMOVE

File Upload empty field: NA

By default, GTTN-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

Analysis lab name	Analysis lab spectra ID	Spectra gatherer	Type of DART TOFMS	Parameter settings	Calibration
Analysis lab Name ▾	Analysis lab Spectra ID ▾	Spectra Gatherer ▾	Type of DART TOFMS ▾	Parameter Settings ▾	Calibration ▾
U.S. National Fish and Wildlife Forensic Laboratory	WD160413	Victor Deklerck (UGent - Woodlab)	DART TOFMS - JEOL (USA)	To be updated	Calibrate woc
U.S. National Fish and Wildlife Forensic Laboratory	WD160408	Victor Deklerck (UGent - Woodlab)	DART TOFMS - JEOL (USA)	To be updated	Calibrate woc
U.S. National Fish and Wildlife Forensic Laboratory	WD160409	Victor Deklerck (UGent - Woodlab)	DART TOFMS - JEOL (USA)	To be updated	Calibrate woc

☐ My file has no header row

DART Reference Raw Data File: *

Allowed file extensions: zip gz tar
gttn_dart.zip REMOVE

To upload DART information, you will need a top-level DART data file, and a compressed DART Raw Data file.

Top-Level DART Data File

The top-level DART file will contain an entry for each sample analyzed in the DART submission, and each entry must include the following information:

- Sample Internal ID or Xylarium ID
- Analysis Lab Name - The lab that performed the DART analysis.
- Analysis Lab Spectra ID




- Spectra Gatherer
- Type of DART TOFMS
- Parameter Settings - The parameter settings on the DART machine used for analysis.
- Calibration Type

Here is an example top-level DART data file:

	A	B	C	D	E	F	G	H
1	Analysis lab name	Analysis lab spectra ID	Spectra gatherer	Type of DART TOFMS	Parameter settings	Calibration type	Internal SampleID	
2	U.S. National Fish and Wildlife Forensic Laboratory	WD160413	Victor Deklerck (UGent - Woodlab)	DART TOFMS - JEOL (USA)	To be updated	Calibrated to series of known wood samples (USFWS)	Tw58	
3	U.S. National Fish and Wildlife Forensic Laboratory	WD160408	Victor Deklerck (UGent - Woodlab)	DART TOFMS - JEOL (USA)	To be updated	Calibrated to series of known wood samples (USFWS)	Tw207	
4	U.S. National Fish and Wildlife Forensic Laboratory	WD160409	Victor Deklerck (UGent - Woodlab)	DART TOFMS - JEOL (USA)	To be updated	Calibrated to series of known wood samples (USFWS)	Tw218	
5	U.S. National Fish and Wildlife Forensic Laboratory	WD160414	Victor Deklerck (UGent - Woodlab)	DART TOFMS - JEOL (USA)	To be updated	Calibrated to series of known wood samples (USFWS)	Tw859	
6	U.S. National Fish and Wildlife Forensic Laboratory	WD160415	Victor Deklerck (UGent - Woodlab)	DART TOFMS - JEOL (USA)	To be updated	Calibrated to series of known wood samples (USFWS)	Tw902	
7	U.S. National Fish and Wildlife Forensic Laboratory	WD160416	Victor Deklerck (UGent - Woodlab)	DART TOFMS - JEOL (USA)	To be updated	Calibrated to series of known wood samples (USFWS)	Tw919	
8	U.S. National Fish and Wildlife Forensic Laboratory	WD160398	Victor Deklerck (UGent - Woodlab)	DART TOFMS - JEOL (USA)	To be updated	Calibrated to series of known wood samples (USFWS)	Tw1018	
9	U.S. National Fish and Wildlife Forensic Laboratory	WD160399	Victor Deklerck (UGent - Woodlab)	DART TOFMS - JEOL (USA)	To be updated	Calibrated to series of known wood samples (USFWS)	Tw1029	
10	U.S. National Fish and Wildlife Forensic Laboratory	WD160400	Victor Deklerck (UGent - Woodlab)	DART TOFMS - JEOL (USA)	To be updated	Calibrated to series of known wood samples (USFWS)	Tw1036	
11	U.S. National Fish and Wildlife Forensic Laboratory	WD160402	Victor Deklerck (UGent - Woodlab)	DART TOFMS - JEOL (USA)	To be updated	Calibrated to series of known wood samples (USFWS)	Tw1040	
12	U.S. National Fish and Wildlife Forensic Laboratory	WD160403	Victor Deklerck (UGent - Woodlab)	DART TOFMS - JEOL (USA)	To be updated	Calibrated to series of known wood samples (USFWS)	Tw1044	
13	U.S. National Fish and Wildlife Forensic Laboratory	WD160404	Victor Deklerck (UGent - Woodlab)	DART TOFMS - JEOL (USA)	To be updated	Calibrated to series of known wood samples (USFWS)	Tw1052	
14	U.S. National Fish and Wildlife Forensic Laboratory	WD160405	Victor Deklerck (UGent - Woodlab)	DART TOFMS - JEOL (USA)	To be updated	Calibrated to series of known wood samples (USFWS)	Tw1085	
15	U.S. National Fish and Wildlife Forensic Laboratory	WD160406	Victor Deklerck (UGent - Woodlab)	DART TOFMS - JEOL (USA)	To be updated	Calibrated to series of known wood samples (USFWS)	Tw1121	
16	U.S. National Fish and Wildlife Forensic Laboratory	WD160410	Victor Deklerck (UGent - Woodlab)	DART TOFMS - JEOL (USA)	To be updated	Calibrated to series of known wood samples (USFWS)	Tw2422	
17	U.S. National Fish and Wildlife Forensic Laboratory	WD160407	Victor Deklerck (UGent - Woodlab)	DART TOFMS - JEOL (USA)	To be updated	Calibrated to series of known wood samples (USFWS)	Tw13504	
18	U.S. National Fish and Wildlife Forensic Laboratory	WD160411	Victor Deklerck (UGent - Woodlab)	DART TOFMS - JEOL (USA)	To be updated	Calibrated to series of known wood samples (USFWS)	Tw26498	
19	U.S. National Fish and Wildlife Forensic Laboratory	WD160412	Victor Deklerck (UGent - Woodlab)	DART TOFMS - JEOL (USA)	To be updated	Calibrated to series of known wood samples (USFWS)	Tw26500	
20								
21								

Compressed DART Raw Data File

The compressed DART Raw Data file should be a .zip, .tar, or .gz file which is a compressed folder of plain text files. Each text file should contain the DART spectra for one sample. Here is an example of the uncompressed folder:

Name	Date modified	Type	Size
 Tw58.txt	9/19/2016 10:59 AM	Text Document	3 KB
 Tw207.txt	9/19/2016 10:58 AM	Text Document	3 KB
 Tw1121.txt	9/19/2016 11:35 AM	Text Document	5 KB

Note that each file name is of the format <sample ID>.txt. This is important for identifying which file is associated with which sample. The file itself should be a collection of weights and peaks, which is the raw DART data:

Milex1.LRP

RT 7.38 - 7.70 (7.79- 8.03) Base=62.6%FS

81.0323	1.4
81.0684	2.5
85.0282	3.8
87.0443	1.2
96.0439	2.1
97.0282	13.4
99.0437	3.7
101.0228	5.9
103.0457	2.6
107.0845	1.8
109.0282	1.0
109.1007	3.7
110.0605	2.4
111.0441	16.3
112.0429	1.8
113.0232	4.7
115.0389	25.2
116.0421	1.2
117.0543	1.4
121.1006	3.0
123.0438	4.3
123.1161	2.0
124.0416	1.2
125.0589	1.1
125.0952	2.8
126.0524	1.6
127.0386	5.0
127.1105	1.2
129.0552	1.7
131.0490	20.1
132.0530	2.0
133.0598	2.6
134.0792	1.2
135.0426	1.1
137.0587	2.5
137.1310	2.1
139.0382	8.2
140.0689	1.0
143.0338	1.1
143.1056	3.0

The file should be in the format:

<Title>

(continues on next page)

(continued from previous page)

```
<DART Configuration>

<weight>\t<peak>
[<weight>\t<peak>...]
```

Isotope

This section collects information about an Isotope analysis:

ISOTOPE REFERENCE DATA INFORMATION

☐ Increment core was used for sampling

Isotope(s) used

☒ ¹³C
☒ ¹⁸O
☐ ¹⁵N
☐ ³⁴S
☐ ⁸⁷Sr
☐ DH

¹³C Isotope standard

¹³C Isotope type

¹⁸O Isotope standard

¹⁸O Isotope type

Isotope Reference Data File: *

Allowed file extensions: txt csv xlsx

File Upload empty field:

By default, GTTN-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

Sample ID	¹³ C	¹⁸ O
Tw58	10	30
Tw207	20	40

☐ My file has no header row

You will be required to indicate which isotopes you used, which isotope standard you used for each isotope, and the type of each isotope (whole wood or cellulose). You will then be required to provide an isotope data file, which contains an entry for each analyzed sample. Each entry must contain a column with the sample ID, and a column with

the measurement for each isotope used. Here is a simple example of an isotope data file:

	A	B	C	D
1	Sample ID	13C	18O	
2	Tw58	10	30	
3	Tw207	20	40	
4				

Genetics

This section collects information about a Genetic analysis:

GENETIC REFERENCE DATA INFORMATION

Marker type(s): *

☒ SNPs
☒ SSRs/cpSSRs
☐ Other

Source of SNPs: * GBS

GBS Type: * ddRad

GBS Sequencer Machine: *

GBS Intermediate Reference File: * - Select -

GBS Alignment File: *

Allowed file extensions: sam bam

Choose File

No file chosen

UPLOAD

VCF File: *

Allowed file extensions: vcf

Choose File

No file chosen

UPLOAD

SSR Machine: *

Ploidy: - Select -

SSRs/cpSSRs Spreadsheet: *

Choose File

No file chosen

UPLOAD

Please upload a spreadsheet containing your SSRs/cpSSRs data. The format of this file is very important! GTTN-TPPS will parse your file based on the ploidy you have selected above. For any ploidy, GTTN-TPPS will assume that the first column of your file is the column that holds the Sample Identifier that matches your sample file.

DNA Quality

For all genetic information, you will need to provide the DNA Quality Score.

The contents of the form in this section vary greatly depending on the type of genetic analysis and the type of genetic markers used.

SNPs

For SNP data, you will first need to identify the source of your SNPs, either GBS, Reference Genome, Transcriptome, or Genotype Assay.

If you selected GBS as your source, you will need to provide the following:

- GBS Type (ddRad, RAD, NextRad, etc)

- GBS Sequencing Instrument name
- GBS Intermediate reference file: either select a reference file from the list of existing reference files on the GTTN Server, or upload your own reference file.
- GBS Alignment file
- VCF File

If you selected Assay as your source, you will need to provide the following:

- Assay Source (MassArray, Illumina, Thermo)
- Assay Design File
- Assay Genotype Table

SSRs/cpSSRs

For microsatellite data, you will need to provide the name of the Sequencing Instrument you used for your analysis, as well as the ploidy of the organism you are analyzing and an SSR spreadsheet containing all of the raw SSR data.

Wood Anatomy

This section collects information about a Wood Anatomy analysis:

ANATOMICAL REFERENCE DATA INFORMATION

ANATOMICAL CHARACTERISTICS

ENTANDROPHRAGMA CYLINDRICUM ANATOMY INFORMATION

Nomenclature:

General:

Vessels:

Tracheids and Fibres:

Axial Parenchyma:

We collect metadata for each species based on the IAWA standards: Nomenclature, General, Vessels, Tracheids and fibres, Axial parenchyma, Rays, Storied structures, Mineral inclusions, Physical and chemical tests.

The screenshot shows a web form titled "MICROSCOPE SLIDES:" in a teal box. Below the title are two orange buttons: "ADD IMAGE" and "REMOVE IMAGE". The form contains two identical sections for adding microscope slides. Each section has a text input field for the image, a "Choose File" button, and an "UPLOAD" button. Above the input fields, it says "Allowed file extensions: img jpg jpeg png svg". Below the input fields, it says "Slide image 1 description:" and "Slide image 2 description:" respectively, followed by a prompt: "Please provide a brief description of the provided image".

You can upload any number of microscope slide images and provide brief descriptions for each.

4.1.7 Submission Review

Finally, once all of the reference and analysis data has been provided, you will be shown a brief summary of the data you are submitting before you mark your submission as complete. This page will contain information provided by the user including the submission name, collection reason, etc. and will allow the user to view previews of the files they have provided to the form. After the user has verified that all of the information is correct, they can click "Submit" and the submission will be sent to administrators for the approval process.

4.2 Reference and Sample Data Search Form

To browse and search reference data that has been uploaded through GTTN-TPPS, we use the reference and sample data search form. The form can be found at gttn.treegenesdb.org/reference.

While the form is still a work in progress, the prototype currently available should be able to give an idea of what might be available in the future.

Currently, there are two possible data types to search by: Reference Data Submissions and Sample Data. We plan to expand this search form so that it is possible for users to search for different data types, filter by different criteria, and access detail views which include the full list of data for elements returned by the search.

4.2.1 Reference Data Submissions

To browse all reference data submissions, select "Reference Data Submission" from the first drop-down menu, then click "Search". If the text field is left blank, then the search results will include every available Data submission that has been approved from GTTN-TPPS:

Reference Data Submission ▼

Project Name ▼

SEARCH

Accession	Project Name	Species	Data Types	Submitting Organization
GTTN-TGDR093	DART Test info	Entandrophragma cylindricum	Sample Data DART Reference Data	test-organization-1
GTTN-TGDR094	Isotope Test info	Afzelia africana	Sample Data Isotope Reference Data	test-organization-1

Here we can see some brief data about the submissions that matched our search.

We can currently filter submissions by project name, species, data type, or submitting organization for example, if we enter something like “DART Test” to the textfield and click search, the form will return only the submissions with the phrase “DART Test” in the project name:

Reference Data Submission ▼

Project Name ▼


DART Test

SEARCH

Accession	Project Name	Species	Data Types	Submitting Organization
GTTN-TGDR093	DART Test info	Entandrophragma cylindricum	Sample Data DART Reference Data	test-organization-1

You can also view details about submissions (as long as you have adequate permissions) from this page by clicking the accession number or the project name. You will then be able to browse the fine details of the submission such as information about individual samples and trees, as well as raw data downloads:

Return to Browse Reference Data



DART DEMO SUBMISSION

Submitted by Peter Richter for University of Connecticut

[Click to show data collection reason](#)
[Download compressed raw files](#)

Species Study Details Trees Samples DART

Accession	GTTN-TGDR021
Project Name	Dart Demo Submission
Submitter	Peter Richter
Species	Entandrophragma utile
Data Type	Sample Data DART Reference Data
File Downloads	dart_accession.xlsx dart_samples.xlsx DART_test.xlsx gttn_dart.zip
Number of Trees	18
Number of Samples	18
Number of DART Reads	3

4.2.2 Sample Data

Similar to the Reference Data Submissions, you can browse all samples by selecting “Sample” from the first drop-down menu, then clicking “Search”:

Sample

Internal Sample ID

SEARCH

1

2

next >

last >

Internal Sample ID	Species	Tissue Type	Source Material
Tw58	Entandrophragma cylindricum	bark	TPPS-DELETE-ME-100
Tw207	Entandrophragma cylindricum	bark	TPPS-DELETE-ME-101
Tw218	Entandrophragma cylindricum	bark	Tw58
Tw859	Entandrophragma cylindricum	bark	Tw207
Tw902	Entandrophragma cylindricum	bark	TPPS-DELETE-ME-100
Tw919	Entandrophragma cylindricum	bark	Tw902
Tw1018	Entandrophragma cylindricum	bark	Tw919
Tw1029	Entandrophragma cylindricum	bark	Tw919
Tw1036	Entandrophragma cylindricum	bark	Tw58
Tw1040	Entandrophragma cylindricum	bark	Tw218
Tw1044	Entandrophragma cylindricum	bark	Tw859
Tw1052	Entandrophragma cylindricum	bark	Tw859
Tw1065	Entandrophragma cylindricum	bark	Tw859
Tw1121	Entandrophragma cylindricum	bark	Tw1040
Tw2422	Entandrophragma cylindricum	bark	Tw1040
Tw13504	Entandrophragma cylindricum	bark	Tw1040
Tw26498	Entandrophragma cylindricum	bark	Tw1040
Tw26500	Entandrophragma cylindricum	bark	Tw1040
Tw58	Afzelia africana	bark	TPPS-DELETE-ME-100
Tw207	Afzelia africana	bark	TPPS-DELETE-ME-101

We can currently filter by Sample ID and species. To search by species, simply select “Species” from the second drop-down menu, and enter something like “Entan” to the textfield and click search, the form will return only the samples which are from a species whose name contains “Entan”:

Sample ▼ Species ▼ Entan

SEARCH

Internal Sample ID	Species	Tissue Type	Source Material
Tw58	Entandrophragma cylindricum	bark	TPPS-DELETE-ME-100
Tw207	Entandrophragma cylindricum	bark	TPPS-DELETE-ME-101
Tw218	Entandrophragma cylindricum	bark	Tw58
Tw859	Entandrophragma cylindricum	bark	Tw207
Tw902	Entandrophragma cylindricum	bark	TPPS-DELETE-ME-100
Tw919	Entandrophragma cylindricum	bark	Tw902
Tw1018	Entandrophragma cylindricum	bark	Tw919
Tw1029	Entandrophragma cylindricum	bark	Tw919
Tw1036	Entandrophragma cylindricum	bark	Tw58
Tw1040	Entandrophragma cylindricum	bark	Tw218
Tw1044	Entandrophragma cylindricum	bark	Tw859
Tw1052	Entandrophragma cylindricum	bark	Tw859
Tw1065	Entandrophragma cylindricum	bark	Tw859
Tw1121	Entandrophragma cylindricum	bark	Tw1040
Tw2422	Entandrophragma cylindricum	bark	Tw1040
Tw13504	Entandrophragma cylindricum	bark	Tw1040
Tw26498	Entandrophragma cylindricum	bark	Tw1040
Tw26500	Entandrophragma cylindricum	bark	Tw1040

CHAPTER 5

Administrator Information
