Plant Ontologies in Forest Trees: TreeGenes

Jill Wegrzyn

Department of Plant Sciences
University of California at Davis
TreeGenes Database

Encompasses Dendrome Resources, DendromePlone, TreeGenes Database & DiversiTree

- Nine modules to store and interrelate data for query and analysis in PostgreSQL and content management through Plone
  - Direct resource for nearly 2,500 forest geneticists representing 800 organizations worldwide. Over 6,000 unique visitors in December 2011.
    - Forest Geneticists Colleague module
    - Literature module
    - Comparative map module
    - Species module
    - Sequencing/Primers module
    - Genotype/SNP module
    - Gbrowse/Genomic
    - Phenotype/Expression module
    - Sample tracking module
Genomic Resources

678 Species Representing 77 Genus

Pinus lambertiana | Sugar Pine

Information | Resources | Map | Publications


NCBI Taxonomy
Completed Genomes:

**Eucalyptus grandis**

**Populus trichocarpa**

**Conifers:**
- **Norway Spruce (Sweden)**
- **White Spruce (Genome Canada)**
- **Douglas fir** (*Pseudotsuga menziesii*), n=13
  - Genome size: 18,700 Mb
  - Genotype to sequence: 412-2
  - Mapping population: 412-2x013-1 (500 F₁ progeny)
- **Sugar pine** (*Pinus lambertiana*), n=12
  - Genome size: 33,500 Mb
  - Genotype to sequence: 6000
  - Mapping population: 5038x5500 (1300 F₁ progeny)
The Forest Tree Genomics Stock Center (FTGSC) at the University of California at Davis provides collection, preservation, and distribution of forest tree DNA to the forest research community. Primary support for this service is provided by the USDA National Institute for Food and Agriculture (NIFA, formerly CSREES) and the USDA Forest Service.

The Forest Tree Genomics Stock Center performs four primary services:
1. Sample Collection & Storage
2. DNA Extraction
3. Web-based Sample Search & Information Retrieval
4. Web-based Sample Ordering

Overview

Sample Submission Form

Before filling out and submitting the form below, after the form has been submitted, you will receive shipping instructions along with a Shipping Receipt which must be included with your shipment. Depending on how you fill out the form below, you may or may not receive barcodes for your samples.

Select your institution or company name:

If not listed, register your lab here.

Contact information:
Principal Investigator’s Name:
Principal Investigator’s E-mail:
Project Coordinator’s Name:
Project Coordinator’s E-mail:
Project Coordinator’s Phone:
Shipping Address:
(For parcel packages)

What is the name of this population set?
Please enter a name for the set of samples being shipped to the FTGSC. For example, a population set may be named “Collotype 14, Study Set 1” or “Smith Lab - 2011 Upstate Samples”. The name entered here will be used to describe your population set within the FTGSC tracking system.

Please describe the population set and the research it has been used for.
Please enter a description for your population set, and provide information about the research studies that have involved these samples. Please feel free to mention prior publications in this description. This description will be used to help other researchers determine whether or not this sample set may be useful for their studies.

Select the tree species:

Select the preservation type:

Select your sample type:

How many individual trees will you be sending samples for?

Please provide the total number of individual (genetically distinct) trees, not the total number of clones.

NOTE: All genetically identical samples MUST share the same FTGSC Barcode.

Have any of these samples been submitted to the FTGSC in the past?
Accurately track samples through collection, DNA extraction, and genotyping

Provide a standard and efficient method to collect and store phenotypic data

Provide a public interface to readily query raw genotype, phenotype, and association results (DiversiTree)

Provide interfaces and database backend to support a DNA distribution center (UCD)
Data Submission

Sample Submission

Phenotypes Defined

<table>
<thead>
<tr>
<th>Phenotype/Eco. Site</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAMUS-AVE-STR</td>
<td>1.59</td>
</tr>
<tr>
<td>TAMUS-BV-4</td>
<td>1.40</td>
</tr>
<tr>
<td>TAMUS-BV</td>
<td>7.50</td>
</tr>
<tr>
<td>TAMUS-BV-V</td>
<td>2.20</td>
</tr>
<tr>
<td>TAMUS-AVE-STR</td>
<td>2.10</td>
</tr>
<tr>
<td>TAMUS-TOT-STR</td>
<td>3.10</td>
</tr>
</tbody>
</table>

Genotypes

<table>
<thead>
<tr>
<th>Accession Number</th>
<th>Phenotype Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>11-1109</td>
<td>AC AC AC AC</td>
</tr>
<tr>
<td>11-111</td>
<td>AC AC AC AC</td>
</tr>
<tr>
<td>11-1110</td>
<td>AC AC AC AC</td>
</tr>
<tr>
<td>11-1140</td>
<td>AC AC AC AC</td>
</tr>
</tbody>
</table>

Genotypes

<table>
<thead>
<tr>
<th>S amples</th>
<th>Raw Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genotypes</td>
<td>Genotypes</td>
</tr>
</tbody>
</table>

Phenotype Data

<table>
<thead>
<tr>
<th>Accession Number</th>
</tr>
</thead>
</table>

DiversiTree

Forest Tree Genetic Stock Center

Sequencing

Literature Object

iPlant Collaborative™

http://dendrome.ucdavis.edu/iplant/
Currently no other repositories to target association data with geo-referenced data
  - dbGAP
  - Dryad
  - Starting with enforcement at the journal level: Tree Genetics and Genomes
The Tree Biology Seed project is an iPlant initiated seed project with the goal to address an iPlant Grand Challenge question. This project intends to address the integration disparity between databases by implementing iPlant's Simple Semantic Web Architecture and Protocol (SSWAP) to make databases interoperable. For the prototype seed project, there will be development to bring in Tree Biology data from genetic, ecological, and physiological data sources. The goal of the project is to create a system to allow Tree Biologists query geographic locales based on geo-referenced coordinates. The Tree Biologists will then be able to explore genetic, physiological, and ecological data sources based on this location to potentially perform an association study by bringing in other data types.

Currently, the seed project plans to prototype access to TreeGenes and Diversitree for genetic data, AmeriFlux for ecological data, and TRY-db for physiological data. Once a prototype system has been implemented, we hope to extend interoperability to Fluxnet, LTER, and NCEAS SIEN for ecological sources. Along these lines, we also hope to extend support to more genomic resources like Genome Database for Rosaceae and Fagaceae Genomics Web.
# SSWAP Ontology

Creating and Contributing to Existing Servlets for Common Genomic Types

## Biology Core Set

Ontologies containing core classes and properties useful for establishing a minimum vocabulary with relations to the Open Biomedical Ontologies and others.

<table>
<thead>
<tr>
<th>OWL Ontology</th>
<th>Ontology Browser</th>
<th>Source File</th>
<th>Ontology Provider</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>genome</td>
<td>Browse genome</td>
<td>genome.owl</td>
<td>SSWAP</td>
<td>Genome-related terms</td>
</tr>
<tr>
<td>map</td>
<td>Browse map</td>
<td>map.owl</td>
<td>SSWAP</td>
<td>Map-related terms</td>
</tr>
<tr>
<td>marker</td>
<td>Browse marker</td>
<td>marker.owl</td>
<td>SSWAP</td>
<td>Marker-related terms</td>
</tr>
<tr>
<td>qtl</td>
<td>Browse qtl</td>
<td>qtl.owl</td>
<td>SSWAP</td>
<td>QTL-related terms</td>
</tr>
<tr>
<td>sequence</td>
<td>Browse sequence</td>
<td>sequence.owl</td>
<td>SSWAP</td>
<td>Sequence-related terms</td>
</tr>
<tr>
<td>sequenceServices</td>
<td>Browse sequenceServices</td>
<td>sequenceServices.owl</td>
<td>SSWAP</td>
<td>Sequence Services’ terms for BLAST, ClustalW, etc.</td>
</tr>
<tr>
<td>taxa</td>
<td>Browse taxa</td>
<td>taxa.owl</td>
<td>SSWAP</td>
<td>Taxa-related terms</td>
</tr>
<tr>
<td>trait</td>
<td>Browse trait</td>
<td>trait.owl</td>
<td>SSWAP</td>
<td>Plant trait-related terms</td>
</tr>
</tbody>
</table>
FLUXNET: A Global Network of Observation Sites
500+ Sites, 10 Regional Networks, 45 Countries

Quantify and understand causes of variation in terrestrial exchange of carbon, water and energy with atmosphere
www.fluxdata.org
TRY – Global Database of Plant Traits

• Scientists compiled three million traits for 69,000 out of the world's ~300,000 plant species.
• Worldwide collaboration of scientists from 106 research institutions
• TRY is hosted at the Max Planck Institute for Biogeochemistry in Jena (Germany)
  – Jointly coordinated with:
    • University of Leipzig (Germany)
    • IMBIV-CONICET (Argentina)
    • Macquarie University (Australia)
    • CNRS and University of Paris-Sud (France)