Cartogratree v3.0: Enabling Forest Tree Genomics through Association Studies

Presented by: Jacob Zieve
Hans Vasquez-Gross, Damian Gessler, David Neale, Jill Wegrzyn

TreeGenes Database
Department of Plant Sciences
University of California at Davis

iPlant Seed Project: Tree Biology Cyberinfrastructure

- **WHO:** Representatives from the Forest Tree Biology Community: Tree Physiologists, Forest Ecologists, Population Geneticists, Bioinformaticians
- **WHY:** Physiological, population genetic, and ecological data lack shared, integrated, geo-referenced data storage
- **HOW:** Cyberinfrastructure to allow data storage, retrieval, integration, and analysis
- **WHAT:** Genetic and genomic data from resequencing, transcriptome, and SNP studies associated with individuals with GPS coordinates. Phenotypes and environmental data with GPS.

- **Database Resources:**
  - Ameriflux (Abiotic Site Data in N. America), WorldClim (Abiotic Data Globally), Try-DB (Biotic/Abiotic Trait Data), TreeGenes (Genomics Data)

- **Tools:**
  - Simple Semantic Web Architecture and Protocol (SSWAP)
  - iPlant Platform
  - iPlant Share
  - iPlant Tree Genes
  - DiversiTree

- **SSWAP benefits:**
  - To accomplish the project goals, web services need to be built to serve the relevant data
  - Most services have a unique ID making integration with existing applications easy. SSWAP semanticizes, reasons, and runs the app+data at the Texas Advanced Computing Center (TACC).
    - Ameriflux, Amplicon, Alignment, Contig, Contig keyword, EST, Multi-Fasta, Primer, Tree Sample, SNP, and Region sample
  - These services have been integrated within DiversiTree by simply allowing users to select records then click a button to bring the data to SSWAP services.
  - Learn more tomorrow 1:30, Meeting House -> Towne room

TreeGenes Database

- Nine modules to store and interrelate data for query and analysis in PostgreSQL
  - Direct resource for nearly 2,500 forest geneticists representing 800 organizations worldwide. Over 8,000 unique visitors in December 2013.
    - 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

DiversiTree Resource

- Advanced Workspace to Facilitate Search and Analysis of Resequencing Data
  - Allows users to search from a variety of starting points
    - Functional Annotations (Candidate Gene)
    - Species
    - Genbank
    - ESTs
    - Sample/DNA-Extraction Data
  - Create customized lists, download data, and view in a variety of formats
    - FASTA format
    - Chromatogram Viewer
    - ACE Format
  - Supports creation of custom SNP lists
    - Raw and normalized SNP scores
    - Genotype, flanking sequence, and alignments
    - Export to Genotyping formats (i.e. Illumina)
DiversiTree Resource
Advanced Workspace to Facilitate Search and Analysis of Resequencing Data

DiversiTree
SSWAP Integration

Selected contigs of interest
Send data to sswap.info

DiversiTree
SSWAP Integration

Built SSWAP pipeline
Display Data Link to CartograTree

DiversiTree + Cartogratreer
SSWAP Integration

Resulting Visualization

CartograTree 3.0
What’s New, Usability

- Simplified, responsive user interface: Bootstrap
- 16,732 markers = ☺. We needed a more powerful backend by leveraging Google Fusion Table layers

- The code: Backbone.js and Slickgrid.js
- Brings a more performant and better user experience
- Access to TreeGenes Data Repository studies

CartograTree
Opening: Layout

TreeGenes
Approximate GPS Marker

Exact GPS Marker

Ameriflux Marker

Try-DB Marker

Once the individual trees of interest are selected, a variety of tools are available including:

- **Common SNP, Common Phenotype, Get WorldClim Data** allow users to view the data in the table or download CSV
- **DiversiTree Input File** allows users to easily explore the data in DiversiTree
- **Common Amplicon** can be used to get sequence data and run a Multiple Sequence Alignment
- **Tree ID Data** brings the tree sample data to the SSWAP pipeline
- **TASSEL** brings the tree sample data to the SSWAP pipeline and leverages iPlant High Performance Computing (HPC) infrastructure to run TASSEL

**TreeGenes Data Repository**

- Allows researchers to submit their study’s data to TreeGenes for archiving
  - The researcher receives a unique accession which can be used in publications (ex. TGDR001)
- The accession holds all submitted information, ranging from genotype, phenotype, environmental, GPS, and meta data
- Researchers can choose to have a Release Date, when the data will be available to the public
- If the data has GPS information and is publicly available, the study will be automatically integrated into CartograTree
CartograTree
TreeGenes Data Repository
dendrome.ucdavis.edu/tgdr

CartograTree
TASSEL: Dataset Selection
You can multi-select also!

CartograTree
TASSEL: Selection in bottom 'Selected Samples' Panel

CartograTree
TASSEL: Selecting and running the TASSEL tool

CartograTree
TASSEL: Sending the data to SSWAP
CartograTree

TASSEL: Getting the results from iPlant’s Discovery Environment

Coming soon...

- User-uploaded phenotypic values, kinship matrices, population structures
- View, sort, filter missing data
- Integration of several map layers (ArcGIS)
- Caveat: Beta

Questions and Comments

Use: dendrome.ucdavis.edu/cartogratree3_beta

Learn More:
Poster TreeGenes: P1008
Poster CartograTree: P1049
Follow us: @TreeGenes

Read More:
DOI: 10.1111/1755-0998.12067

Acknowledgements

TreeGenes:
Jill Wegrzyn
David Neale
Jacob Zieve
Hans Vasquez-Gross

iPlant Collaborative:
Damian Gessler
Martha Narro

TRY:
Jens Kattge
Gerhard Borkent

WorldClim:
Robert Hijmans

User:
Robert H. Hijmans