Case Study: Enabling Forest Tree Genomics through Association Studies with the use of SSWAP & CartograTree

Presented by: Hans Vasquez-Gross
Jacob Zieve, Damian Gessler, David Neale, Jill Wegrzyn
TreeGenes Database
Department of Plant Sciences
University of California at Davis

SSWAP Talk
Overview

• Who, Why, and What summary of the project
• TreeGenes Introduction
• SSWAP Services
  — Built SSWAP Services
• SSWAP Integration within existing application
  — DiversiTree
• SSWAP Integration within new application
  — CartograTree

iPlant Seed Project: Tree Biology Cyberinfrastructure

• WHO?: Representatives from the Forest Tree Biology Community: Tree Physiologists, Forest Ecologists, 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
• Map-based tool to visualize the geo-referenced data
• 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)

SSWAP Services

What is available

• To accomplish the project goals, web services need to be built to serve the relevant data
• Most services operate on a unique identifier, lending itself well to being integrated into existing applications
  — 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.

TreeGenes Database

Encapsulates DendroCenter Resources, DendroCenterDB, TreeGenes Database & DiversiTree

• 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/Primer 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

CartograTree 3.0
What’s New
- Simplified user interface
- With ~17,000 map markers, we needed a more powerful backend by leveraging Google Fusion Table layers
- Through implementation of Backbone and Bootstrap
- Brings a more performant and better user experience
- Access to TreeGenes Data Repository studies

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

- Common Amplicon can be used to get sequence data and run a Multiple Sequence Alignment
- 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
- 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

CartograTree 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

TASSEL: Dataset Selection

Select samples in the 'Selected Samples' Panel

TASSEL: Selecting and running the TASSEL tool

TASSEL: Sending the data to SSWAP

TASSEL: Getting the results from iPlant's Discovery Environment
Future plans...

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

Acknowledgements

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

iPlant Collaborative: Damian Geissler, Martha Narro

TRY: Jens Kattge, Gerhard Boerisch

WorldClim: Robert Hijmans

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