

# Bioinformatic Solutions for Data Storage, Analysis, and Interpretation in Forest Genomics: Overview of the Tools and Resources from the Dendrome Project

CATTAGCTCTGGTCATCAAGTCATCCATGATTAGCT



Jill L. Wegrzyn, Ben Figueroa, Minyoung Choi, John D. Liechty, Andrew J. Eckert, David B. Neale

<http://dendrome.ucdavis.edu>



## Phenotype = Genotype + Environment



**Adaptation**



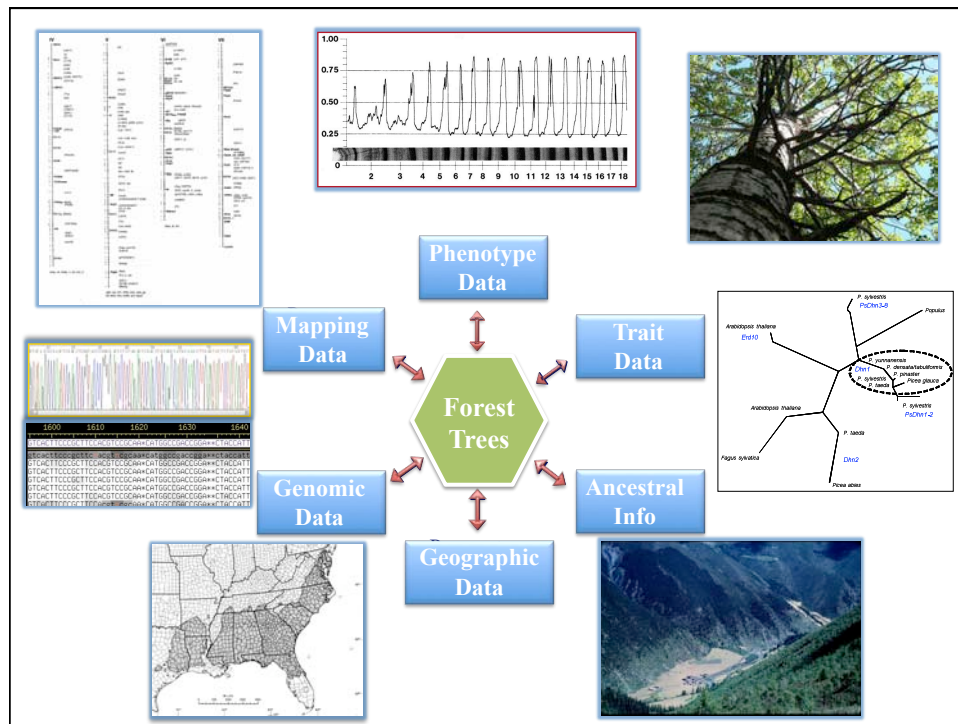
**Breeding**



**Conservation**

<http://dendrome.ucdavis.edu>





## What Makes a Database a Good Resource?

- **Integration of Public Information (Repositories)**
  - Hand-selected for organisms of interest
- **Data from the Community!**
  - Clear pathways for users to submit data
- **Flexibility**
  - Data is generated at increasing rates
  - New technologies come with new standards, metadata, and new storage requirements
- **Integration (Biggest Challenge)**
  - Exchange of data with other plant genomics databases and NCBI
  - Working towards standards (Ontologies and nomenclatures)
  - Exchange data within the same database
    - i.e. Find a particular marker on multiple genetic maps
- **Regular and accurate curation**
  - Knowledgeable curators and community level curation
- **Interfaces**
  - Extensive research into what users want to achieve
  - Optimize the relevant queries
  - Easy to use!

## Interacting with Public Databases

- **Genbank/DDJB/TIGR/EMBL (Sequence)**
  - dbEST
  - dbSNP
  - SRA
  - Trace Archive
  - PopSet
- **Expression**
  - GEO
- **Proteomics**
  - SwissProt/TrEMBL
  - Peptidome
  - PDB
  - PIR
- **Metabolomics**
  - MetaCyc
  - KEGG

<http://dendrome.ucdavis.edu>

Dendrome



### GDR | Genome Database for Rosaceae

Login

Home
General Info
Species
Projects
Maps
Search
Tools
Community
Calendar
Contact

### Genomic, Genetic, and Breeding Resources for Rosaceae Crop Improvement

### Peach Genome Sequence Released



#### Welcome to the Genome Database for Rosaceae


The Genome Database for Rosaceae (GDR) is a curated and integrated web-based relational database providing centralized access to Rosaceae genomics and genetics data and analysis tools to facilitate cross-species utilization of data. Supported by the NSF Plant Genome Program from 2003 to 2008, GDR is now funded until 2013 by the USDA Specialty Crop Research Initiative. As part of tree fruit Genome Database Resources (tFGDR), GDR will be expanded to include whole genome sequences and annotations for apple, peach, and strawberry, transcript data, metacyc pathways, large scale phenotype and genotype data, breeding data, controlled vocabularies, and new analysis tools, all of which will be implemented within the open source infrastructure of drupal and chado (tripal).

#### Whats new in GDR?

- The peach genome v1.0. View and download the predicted genes, genome sequences, putative protein and EST homologs, Rosaceae genera mapped transcripts, mapped markers, and blast your sequences against the datasets.
- Five new maps are available in CMap for prunus, almond and apple.
- GDR redesigned in Drupal. Any problems/comments? Please let us know!
- Batch BLAST server implemented, compare batches of sequences against available databases and get parsed results in excel.
- Report a problem | Ask us a question | Post a job | Post a meeting or event | What's been fixed in GDR? | What are we working on? | Used GDR data or tools in your research? how to reference us.

Community News (8)

- WSU Press Release about Peach Genome Sequence
- The peach genome sequence is released !
- Fruit and Nut Workshop held Saturday, January 9, 8am-12.30pm at PAG 2010
- Highlighted Publication - "Development and Bin Mapping of a Rosaceae Conserved Ortholog Set (COS) of Markers"
- RosEXEC meeting held Sunday, January 10, 8am-10pm in the Towne Room at PAG 2010
- Ksenija Gasic, Amy Iezzoni, Jay Norelli, John Clark, Kevin Folta, David Chagne and Herman Silva elected to RosEXEC.
- Rosaceae Community RosBREED project funded by the 2009 Specialty Crop Research Initiative
- more announcements ...



# Dendrome

A Forest Tree Genome Database


Welcome
Research
TreeGenes
Resources
Events
Employment
Links
Help

**Welcome to the Dendrome Project!**

Dendrome is a collection of forest tree genome databases and other forest genetic information resources for the international forest genetics community. Dendrome is part of a larger collaborative effort to construct genome databases for major crop and forest species.

**International Conifer Genome Network (CGN)**  
The Conifer Genomics Network (CGN) is a virtual nexus for researchers working in conifer genomics. The goal of the CGN is to facilitate information exchange among researchers throughout the world and to serve as a forum for advancing conifer genome sciences. The CGN web site is maintained by the Dendrome Project at the University of California, Davis.

Learn more about Member Organizations  
Join CGN



**HEADLINES ON FOREST TREES**

- New 'Tree of Life' established for one of the largest groups of bacteria  
A new 'tree of life' has been constructed by researchers for the gamma-proteobacteria, a large group...
- Clash of the mites: Hot on the heels of a destructive coconut pest  
Biological control experts are sending mites after their own kind as researchers make headway in an ...
- NASA, Google data show North Korea logging in protected area  
Using NASA satellite data and Google Earth, a Purdue University researcher has reported finding evid...

**UPCOMING EVENTS**

International Symposium on Biology of Rare and Endemic Plant Species  
Muğla-Turkey  
May 26-29, 2010

High Five Conference: The Future of High-Elevation Five-Needle White Pines in Western North America  
Missoula, MT  
June 28-30, 2010

**UPDATES**

Project		Genomic Data	Species
<b><u>ADEPT2 – Allele Discovery of Economic Pine Traits 2</u></b>		-Resequencing -Primer Resource -Genotyping -Phenotyping	Pinus Taeda
<b><u>DFGP – Douglas-fir Genome Project</u></b>		-EST Sequences -Resequencing -Genetic Maps -Phenotyping	Pseudotsuga menziesii
<b><u>WHISP – White Pines Resequencing Project</u></b>		-Resequencing -Genotyping -Phylogenetic Data	32 soft pines
<b><u>CRSP - Comparative Resequencing in Pinaceae</u></b>		-Resequencing -Genotyping	Pinus elliottii, Pinus radiata, Pinus sylvestris, Pinus pinaster, Pinus lambertiana, Pinus abies, Pseudotsuga menziesii
<b><u>ACE-SAP - Alpine Ecosystems in Changing Environments: Biodiversity Sensitivity and Adaptive Potential</u></b>		-Resequencing -Genotyping -Geographic	Pinus mugo, Pinus abies, Pinus cembra, Larix decidua, Abies alba
<b><u>PBGP – Poplar Biofuels Genome Project</u></b>		-Resequencing -Genotyping -Phenotyping	Populus trichocarpa, Populus nigra, Populus Fremontii

**TreeGenes Overview**  
Welcome to the TreeGenes database! TreeGenes is a comparative mapping database...  
**Species Directory**  
Species of the week: Jeffrey Pine (*Pinus jeffreyi*)  
**Literature Directory**  
Recent Publications:  
1. Gonzalez-Martinez, Santiago C. et al. *Molecular Biology and Evolution* 2008 25:4377-4388.  
2. Gonzalez-Martinez, Santiago C. et al. *Plant Molecular Biology* 2008 64:465-474.  
3. Ruiz-De-la-Herrera, Beatriz et al. *Molecular Biology and Evolution* 2008 25:465-474.  
4. Tringali, D. et al. *Molecular Biology and Evolution* 2008 25:3617-3630.  
5. Hernandez, Miguel et al. *Molecular Biology and Evolution* 2008 25:2892-2908.

**TreeGenes:Scripts**  
Here are a few resources available to facilitate bioinformatic needs...  
addcolumn.pl  
getcolumns.pl  
grepseq.pl  
gseq.pl

**Dendrome**  
Welcome to Dendrome!  
Dendrome is a collection of forest tree genome databases and other forest genetic information resources for the international forest genetics community.

<http://dendrome.ucdavis.edu>

Dendrome Project and TreeGenes Database

<http://dendrome.ucdavis.edu>

Dendrome

## History of the Dendrome Project

- AceDB distributed database of curated genetic maps
  - Completely manually curated
  - Included all Forest Trees (Pinus, Eucalyptus, Populus, Quercus)
- First web resources developed around conifers
  - Focus on genetic maps (comparative: Gramene's CMAP) and ESTs
  - Oracle database with web interfaces for access
- Development of more extensive & flexible tools to deal with high-throughput projects
  - Large-scale Resequencing Projects (DiversiTree)
  - Interfaces to allow users to directly participate in the curation of the data
- Expansion of TreeGenes to integrate with all Forest Trees, Public Repositories
  - Exchange of data with other plant genomics databases and NCBI
- Greater focus on the goal of connecting genotype and phenotype
  - Development of interfaces to maintain LIMS system (track samples and phenotypes)
  - Store genotype information and allow for large-scale association studies
  - Preparation of DNA Inventory goal

<http://dendrome.ucdavis.edu>



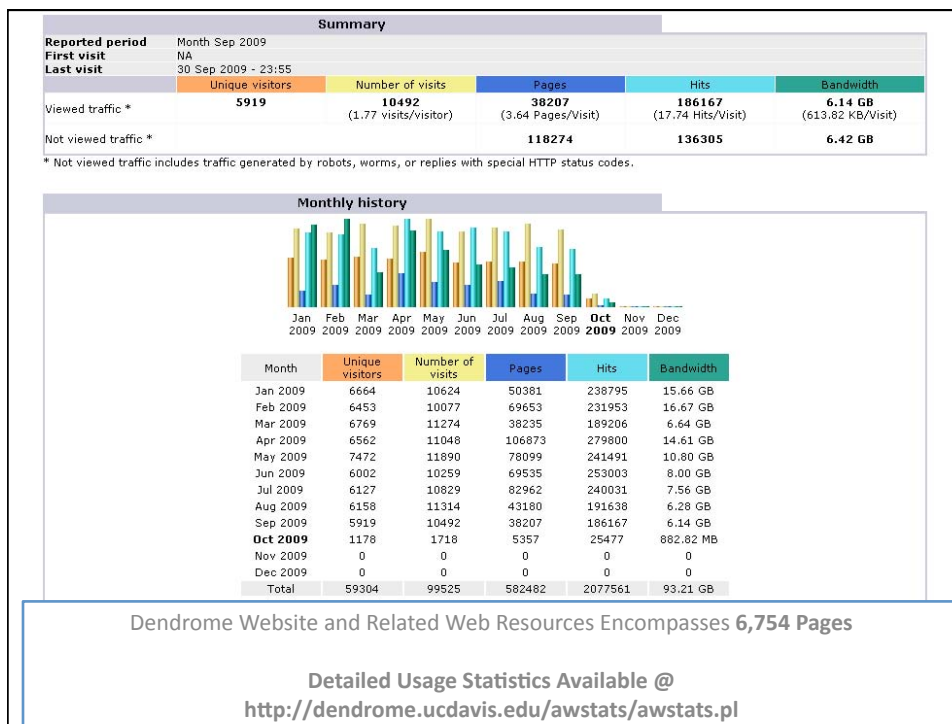
## Dendrome Project


Encompasses Dendrome Resources, DendromePlone, TreeGenes Database & DiversiTree

- Community Resources Featured by Dendrome Pages
  - Links Directory
    - Collaborative community resource to document online resources
  - Local BLAST utilities
    - NCBI BLAST and Local database resources updated regularly
  - Tools and Scripts
    - Extensive documentation on software developed in-house and publicly available
  - News and Events in the Forest Genetics community
  - Portal to TreeGenes Database
  - Portal to DiversiTree Interface
  - Portal to TreeGenes Plone
  - Portal to outreach resources

<http://dendrome.ucdavis.edu>








**Dendrome**  
A Forest Tree Genome Database

Welcome
Research
TreeGenes
Resources
Events
Employment
Links
Help

Welcome to the Dendrome Project!



**Resources**

- Scripts
- BLAST
- FASTA
- Tools
- Databases
- TreeGenes
- Plant Ontology

**Updates**

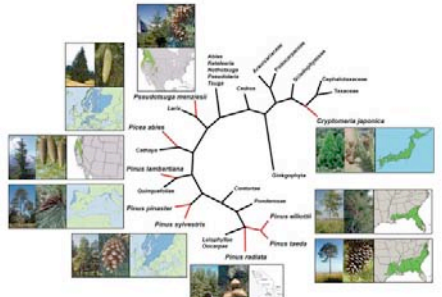
- New EST analysis and submission pipeline available for use!
- Plant Gene Ontology database ported into mysql

**Links**

- Conifer Genome Network
- Conifer Genome Project
- TreeGenes Database

**Tools and Resources**

Welcome to the Resources & Support pages at TreeGenes. The following pages host a wide variety of resources for researchers in the conifer community. Details regarding local databases hosted on our servers as well as custom repositories can be found here. The majority of tools described here are available via web interface to users currently holding an account. All scripts described here are available for use by the community.



**Events**

- Population Genetics and Genomics of Forest Trees: From Gene Function to Evolutionary Dynamics and Conservation  
Madrid, Spain  
October 18 - 19, 2009
- BEIRO Division 2 Joint Meeting  
Marina Pine Beach Resort, Antalya, Turkey  
October 9-13, 2009
- Plant and Animal Genome XV Conference  
Town and Country Convention Center, San Diego, California  
January 13-17, 2007

**UPDATES**

New EST analysis and submission pipeline available for use!

Plant Gene Ontology database ported into mysql

New Forestry Careers and Education Outreach Website is Live!

**Links**

Conifer Genome Network

Conifer Genome Project

TreeGenes Database

Dendrome Phone

Neale Lab

Forestry Careers and Education Resource

**Links :: Plant Model Organisms**

**Plant Model Organisms :: General Resources**

- Canada's Forest Network  
The Canadian Wood Council CWC is the national association representing Canadian manufacturers of wood products used in construction, CWC works with building codes and standards and by developing and communicating technical information to architects, engineers and builders
- PlantGDB || View Abstracts  
Resources for Plant Comparative Genomics such as sequence download, annotation, and comparison tools
- TreeBase || View Abstracts  
A relational database designed to manage and explore information on phylogenetic relationships
- USDA Plants Species Database  
The PLANTS Database provides standardized information about the vascular plants, mosses, liverworts, hornworts, and lichens of the U.S. and its territories.
- Gymnosperm Database  
Basic information about gymnosperms for the public. Taxonomic tree
- Native Conifers of North America  
Pretty pictures of trees and how to identify
- American Conifer Society Database  
Community dedicated to conifers with web links, photos and forums.
- Plants for a Future Database (Pinaceae)  
Database of plants with information about habitat and uses (website sometimes goes down at end of month)
- BoDD - Botanical Dermalogy Database  
Summary of plants in Pine family and where they are found
- ITIS - Integrated Taxonomic Information System  
Government website with authoritative taxonomic information on plants, animals, fungi, and microbes of North America and the world.
- MPSPlantDB - MPS Plants Databases || View Abstracts  
The MPS plant genomics group focuses on the bioinformatics of plant genomes.
- Plant Transcription Factor Database || View Abstracts  
PlnTFDB is a public database arising from efforts to identify and catalogue all Plant genes involved in transcriptional control.
- ChromDB || View Abstracts  
Displays chromatin-associated proteins, including RNAi-associated proteins, for a broad range of organisms. Our primary focus is to display sets of highly curated plant genes predicted to encode proteins associated with chromatin remodeling.
- TIGR Plant Genomics Group  
TIGR's Genome Projects are a collection of curated databases containing DNA and protein sequence, gene expression, cellular role, protein family, and taxonomic data for microbes, plants and humans.
- Plant MPS Databases || View Abstracts  
Good starting point when working with plants with sequence downloads

**Plant Model Organisms :: Fruits**

- GDR (Genome Database for Rosaceae) || View Abstracts

**Plant Model Organisms :: Arabidopsis**

- GABI (Genome Analysis of the Plant Biological System)
- Gramene || View Abstracts  
A resource for comparative grass genomics
- The Arabidopsis Information Resource || View Abstracts

**EVENTS**

Opportunities, challenges and limitations of genomic-based technologies in forest tree breeding and forest genetics


Freiburg, Germany  
October 7-9, 2009

Genomics of Forest and Ecosystem Health in the Fagaceae (Beech Family)

Research Triangle Park, North Carolina  
November 10-13, 2009

International Symposium on Biology of Rare and Endemic Plant Species


Mugla-Turkey  
May 26-29, 2010



**People**


David Neale  
Principal Investigator

Jill Wegrzyn  
Bioinformaticist



**Welcome** | **Research** | **TreeGenes** | **Resources** | **Events** | **Employment** | **Links** | **Help**

**Welcome to the Dendrome Project!**



**Help**

Help Desk

New Account

Tutorials

FAQs

Forum

**Updates**

New EST analysis and submission pipeline available for use!

Plant Gene Ontology database ported into mysql

**Links**

Conifer Genome Network

Conifer Genome Project

**Tutorials**

- Computer Related
  - Multiple Sequence Alignments
  - Sequence Similarity Searching
  - Sequence Analysis

**Computer Related**

- General CVS Tutorial
- Getting Started with Unix Commands
- Installing Cygwin/X

**Multiple Sequence Alignments**

- Using Boxshade
- Using Multiclustal
- Using ClustalX

**Sequence Similarity Searching**

- Sequence Similarity Searching: Understanding and Using Web Based BLAST
- Using Command Line BLAST

**Sequence Analysis**

- Searching for Patterns in Protein Sequences

**Events**

IUFRO Division 2 Joint Meeting:


Maritime Pine Beach Resort, Antalya, Turkey  
October 9-13, 2006

Plant and Animal Genome XV Conference

Town and Country Convention Center, San Diego, California  
January 13-17, 2007

Tree Biotechnology 2007

Ponta Delgada, Azores, Portugal  
June 3-8, 2007





### TreeGenes::Scripts

Here are a few resources available to facilitate bioinformatic needs. All of the scripts here are written in Perl and require libraries that have been installed on our system. These utilities assist with parsing, processing, and sequence comparisons. You may add a Perl script that you have developed here. If you encounter difficulties with any of the scripts listed here, please contact the help desk.

**addcolumn.pl**

This script adds a column to a column file specified by an arithmetic string  
[ more info.. ]

**getcolumns.pl**

This script extracts a column from a multicolumn file (splits on whitespace by default). You can split more than one column at a time by - or -  
[ more info.. ]

**grepseq.pl**

Extract sub-sequences from sequences on stdin based on a (perl) regular expression given on the command line. Input sequences in labeled fasta format. By default the labels are searched using the regex.  
[ more info.. ]

**gseq.pl**

batchextract retrieves one or more sequence entries from NCBI specified by accession numbers. The script can be given either as standard input, as separate arguments or from a file.  
[ more info.. ]

**histogram.pl**

Makes a 2D or 3D histogram data-set for gnuplot from data in specified column(s) of an input file, starting with a real number are ignored. It also understands "framed" tables dumped from databases.  
[ more info.. ]

**mmaquery.pl**

mmaquery.pl retrieves the accession numbers of all the mRNA sequences for a specified organism from Genbank, omitting ESTs, STSs, working drafts and patents. If the optional third argument is not given, the output is written to STDOUT

#### Events

- Plant and Animal Genome XV Conference**  
Town and Country Convention Center, San Diego, California  
January 13-17, 2007
- Tree Biotechnology 2007**  
Ponte Delgada, Azores, Portugal  
June 3-8, 2007
- Western Forest Genetics Association (WFGA)**

#### TreeGenes::Scripts - blast\_summary\_tophits.pl

**blast\_summary\_tophits.pl**

This script will parse a blast or blastn result and has two output formats that summarize the blast output, see below for example, both outputs ignore Queries with no hits

<b>Location:</b>	/usr/local/genome/bin
<b>Usage:</b>	blast_summary_tophits.pl

**Input:**

Sample blast output below:  
BLASTN 2.2.12 [Aug-07-2005]

Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.* 25:3389-3402.

Query= 0\_10010(1354931) SuperScript+2009 (713 letters)

Database: /common/data/ChloroplastDB\_nt.fasta  
9728 sequences; 7,802,122 total letters

**Output:**

- 1) It can parse the BLAST results and summarize in the following tab delimited format:
- 2) or the script can take the above information and output for each queryID the following information in tab delimited format:  
<#ofuniques> <#ofalignments>  
lowest evalue has highest priority, then if there is a tie for e-value, highest identity has the next highest priority, if tie

### Databases

#### Protein Databases

Database	Description	Location	Help
nr	All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF	TG/LOB	[more]
SwissProt	Swissprot protein sequence database	TG/LOB	[more]
Aradipopsis	Non-redundant Arabidopsis protein sequences from GenBank	TG/LOB	[more]
TAIR6	Fasta formatted protein sequences from TAIR	TG/LOB	
RefSeq Protein	Refseq peptide sequences	TG/LOB	
Uniprot/SwissProt Plants	Non-redundant Curated Plant protein sequences	TG/LOB	

#### Nucleotide Databases

Database	Description	Location
RefSeq Genomic	RefSeq nucleotide sequences	TG/LOB
Arabidopsis BAC Genomic	Arabidopsis BAC Genomic from Genbank	LOB
Arabidopsis BAC EST Genomic	Arabidopsis BAC EST Genomic from Genbank	LOB
Arabidopsis Nucleotide	Arabidopsis Genome nucleotide from Genbank	TG/LOB
TAIR CDS	Arabidopsis CDS from TAIR	TG/LOB
TAIR Sequences	Non-redundant Arabidopsis from TAIR	TG/LOB
EST others	All EST sequences except human and mouse	TG/LOB

#### 3. Tree Databases

Database	Description	Location
Pseudotsuga Menziesii	Fasta formatted nucleotide database	TG/LOB
Pinus taeda	Fasta formatted nucleotide database	TG/LOB
Pinus glauca	Fasta formatted nucleotide database	TG/LOB
Pinus sitchensis	Fasta formatted nucleotide database	TG/LOB
Populus balsamifera	Fasta formatted nucleotide database	TG/LOB
Populus tremula x Populus tremuloides	Fasta formatted nucleotide database	TG/LOB

#### Basic BLAST

Use BLAST by pasting in a nucleotide or amino acid sequence in FASTA format, or simply using an accession or GI number.

- nucleotide blast - Search a nucleotide database using a nucleotide query
  - Algorithms: blastn, megablast, discontinuous megablast
- protein blast - Search protein database using a protein query
  - Algorithms: blastp, psi-blast, phi-blast
- blastx - Search protein database using a translated nucleotide query
- tblastn - Search translated nucleotide database using a protein query
- tblastx - Search translated nucleotide database using a translated nucleotide query

[Use BLAST]

#### PSI-BLAST

(PSI) Position Specific Iterated BLAST - Search databases of sequence profiles to find members of the same protein family. First round runs BLAST and constructs a position-specific score matrix (PSSM) from significant hits while iterations use this PSSM as the query on profile databases. Uncovers more distant protein relationships.

[Use PSI-BLAST]

#### PHI-BLAST

(PHI) Pattern Hit Initiated BLAST - Identify proteins that preserve a specified pattern contained within your query. Query includes a protein sequence and a regular expression defining the pattern of interest.

[Use PHI-BLAST]

#### BLAST 2 Sequences

Upload or paste 2 sequences and BLAST them against each other.

[Use BLAST 2 Sequences]

## BLAST: Sequence Similarity Searches

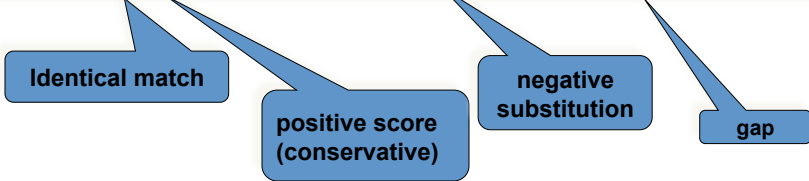
## BLAST Output: Alignments

```

>gi|127552|sp|P23367|MUTL_ECOLI DNA mismatch repair protein mutL
Length = 615

Score = 42.0 bits (97), Expect = 3e-04
Identities = 26/59 (44%), Positives = 33/59 (55%), Gaps = 9/59 (15%)

Query 9  LPKNTHPFLYLSLEISPQNVVDVNVHPTKHEVHF-----LHE---ESILEV-QQHIESKL 58
      L + P L LEI P VDVNVHP KHEV F +H+ + +L V QQ +E+ L
Sbjct 280 LGADQQPAFVLYLEIDPHQVDVNVHHPAKHEVRFHQSRVLVHDFIYQGVLSVLQQQLETP 338
    
```



**Resources**

Scripts

BLAST

FASTA

Tools

Databases

TreeGenes

Plant Ontology

**Updates**

Now EST analysis and submission pipeline available for use!

Plant Gene Ontology database ported into mysql!

**Links**

Conifer Genome Network

Conifer Genome Project

TreeGenes Database

Neale Lab

### TreeGenes :: Tools and Resources

The Tools page contains a list of the currently available bioinformatics software packages installed on the TreeGenes server as well as the Lobioly Server. The majority of these tools are available through a web interface and those that are not can be accessed through the command line

Tool	Description	Version	Location	Manual	Help	Online
bioperl	Bioperl is a collection of perl libraries that ease common bioinformatic needs and integrate with common file formats	1.4	TreeGenes	[more]	[help]	
bl2seq	bl2seq is part of NCBI's BLAST package and allows alignment between 2 sequences	2.2.10	TreeGenes/Lobioly	[more]	[help]	X
blastall	blastall is part of NCBI's BLAST package and uses all BLAST implementations: blastp, blastn, blastx, tblastx, and tblastn	2.2.10	TreeGenes/Lobioly	[more]	[help]	X
blastclust	blastclust clusters both protein and DNA sequences based on resulting BLAST alignments	2.2.10	TreeGenes/Lobioly	[more]	[help]	X
blastpgp	blastpgp is part of NCBI's package and is intended for gapped blast searches	2.2.10	TreeGenes/Lobioly	[more]	[help]	X
boxshade	BOXSHADE provides display options for multiple alignments from other tools	3.3.1	TreeGenes	[more]	[help]	X
cap3	CAP 3 is a contig assembly program that can take in FASTA formatted sequences and quality scores	1.0	TreeGenes	[more]	[help]	
clustalw	ClustalW is a multiple alignment utility for DNA and protein sequences	3.3.1	TreeGenes	[more]	[help]	
consed	Consed is a graphical interface for editing phrap output.	13.0	TreeGenes	[more]	[help]	
fbat	FBAT implements a broad class of Family Based Association Tests, adjusted for population admixture.	1.5.5	TreeGenes	[more]	[help]	
formatdb	Formatdb allows custom databases to be formatted for use with NCBI's BLAST suite.	2.2.10	TreeGenes	[more]	[help]	
genscan	Genscan is a gene identification program for applications in both vertebrates and invertebrates.	1.0	TreeGenes	[more]	[help]	X
gold	GOLD (graphical overview of linkage disequilibrium) is a tool for the graphical summary of linkage disequilibrium in genetic data. The summary is well suited to the analysis of dense genetic maps, where contingency tables are cumbersome to interpret.	1.1.0	TreeGenes	[more]	[help]	
haplo	Haplo uses the EM algorithm to estimate the frequencies of multi-site haplotypes.	0.7	TreeGenes	[more]	[help]	
hmmer	HMHER produces multiple sequences alignments for the purpose of database queries and to identify homologues	2.3.2	TreeGenes/Lobioly	[more]	[help]	

**Events**

Plant and Animal Genome XV Conference


Town and Country Convention Center, San Diego, California  
January 13-17, 2007

Tree Biotechnology 2007

Ponta Delgada, Azores, Portugal  
June 3-8, 2007

Western Forest Genetics Association (WFGA)

Galveston, Texas  
June 19-27, 2007



**People**

David Neale  
Principal Investigator

Jennifer Lee  
Bioinformaticist

Jill Wagryn  
Bioinformaticist


Brandon Terase  
Developer

## TreeGenes Database

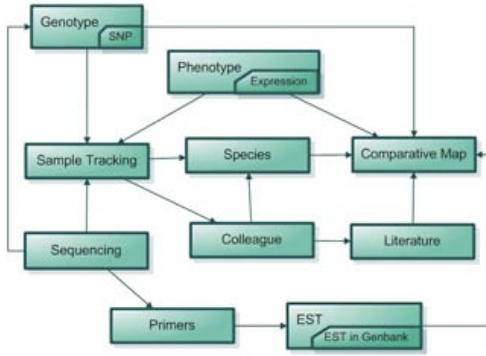
Encompasses Dendrome Resources, DendromePhone, TreeGenes Database & DiversiTree

- Ten modules to store and interrelate data for query and analysis in PostgreSQL
  - Direct resource for nearly 2,500 forest geneticists representing 800 organizations worldwide
    - Forest Geneticists Colleague module
    - Literature module
    - EST annotation pipeline and module
    - Comparative map module
    - Species module
    - Sequencing module
    - Primers module
    - Genotype/EST module
    - Phenotype/Expression module
    - Sample tracking module

http://dendrome.ucdavis.edu



# TreeGenes - Modular Organization



<http://dendrome.ucdavis.edu>



### Literature Directory

Recent Publications

1. Jia, Xiaoyun X, Wang, Wang-Ya W, Ren, Ligang L, Chen, Qi-Jun Q, Mendi, Venugopal V, Wilcutt, Benjamin B, Dirksen, Randy R, Tang, Xiaoping X, Tang, Gullang G. DIFFERENTIAL AND DYNAMIC REGULATION OF MIR398 IN RESPONSE TO ABA AND SALT STRESS IN *POPULUS TREMULA* AND *ABRUSOPSIS THALIANA*. *Plant molecular biology* 2009 71:51-59.
2. Sarfraz, M, A. Estrella MA, Rodon, Rodrigo R, Valero, M A, José MA, Aragón, Mónica M, Valledor, Luis L, Rodríguez, José L, J. Torop, Petar E, Peñalva, M A, Jesús MA, Rodríguez, Roberto R. ACETYLATED H4 HISTONE AND GENOMIC DNA METHYLATION PATTERNS DURING BUD SET AND BUD BURST IN *GASTANIA SATIVA*. *Journal of plant physiology* 2009 166:1360-1369.
3. Ebrahimi, A., Zare, A., Fatahi, B., Vamankhasti, M., Ghasemi. STUDY ON SOME MORPHOLOGICAL AND PHYSICAL ATTRIBUTES OF WALNUT USED IN MASS MODELS. *Scientia Horticulturae (Amsterdam)* 2009 121:490-494.
4. Aerts, Ren, Calaghan, Terry V., Dorcasal, Ellen, Van Logtenhove, Richard S. P., Cornelissen, Johannes H. C. SEASONAL CLIMATE MANIPULATIONS RESULT IN SE LEVELS AND ISOTOPIC COMPOSITION IN A SUB-ARCTIC PLANT REGENERATION IN STONE PINE (*PINUS PINE TISSUE AND ORGAN CULTURE* 2009 96:165-179).

Search:

Author:  Submit  [Advanced Query]

Browse: A - B - C - D - E - F - G - H - I - J - K - L - M - N - O - P - Q - R - S - T - U - V - W - X - Y - Z

The Dendrome project maintains a comprehensive literature database is updated bi-weekly with bibliographic information. The 15452 journal articles found here relate to forest tree genetics and are acquired from a multitude of public repositories.

Search Results

Quang, H.D.; Sada, S.; Henda, H. PATTERN OF NUCLEOTIDE DIVERSITY AT THE METHIONINE SYNTHASE LOCUS IN FRAGMENTED AND CONTIGUOUS POPULATIONS OF A WIND-POLLINATED TREE, *QUERCUS MONGOLICA* VAR. *CRISPULA*. *The Journal of Heredity*, 2005, Vol. 96, No. 1, pp. 1-10.

Reperbas, Marjorie; Scriver, Alexander G.; Huser, Herman P. A.; Allen-B, Peter; Conover, Lu J.; Brown, Leon A. THERMAL ENERGY DISSIPATION AND ITS COMPONENTS IN TWO DEVELOPMENTAL STAGES OF A SHADE-TOLERANT SPECIES, *NOTODAPHNE NETTIDA*, AND A NOT SHADE-TOLERANT SPECIES, *NOTODAPHNE NETTIDA*. *Tree Physiology*, 2009, Vol. 29, No. 1, pp. 1-12.

Olivares, M. V.; Iqbal, E. V. GENETIC CONSEQUENCES OF IRRADIATION IN A SCOTS PINE (*PINUS SCOTICA*). *Journal of Heredity*, 2005, Vol. 96, No. 1, pp. 1-10.

- o Currently referencing **15,452 journal articles**
- o Literature updated monthly from PubMed and BIOSIS
- o Fully connected to species, colleague, and related genomic data sets (genetic maps, ESTs, proteins...)

### Literature Detail

Quang, H.D.; Sada, S.; Henda, H. PATTERN OF NUCLEOTIDE DIVERSITY AT THE METHIONINE SYNTHASE LOCUS IN FRAGMENTED AND CONTIGUOUS POPULATIONS OF A WIND-POLLINATED TREE, *QUERCUS MONGOLICA* VAR. *CRISPULA*.

Genetic variation is usually high within populations, and differentiation is usually low among populations of wind-pollinated outcrossing trees. In a recent population contraction causes little change in the degree of genetic diversity and differentiation among populations. The aim of this work was to determine whether or not a recent population decline had influenced the allele frequency spectrum and association among variants of different sites on the nuclear house-keeping locus methionine synthase (370-1402 bp in length) in the oak species *Quercus mongolica* var. *crispula*. A total of 522 sequences from 19 populations were randomly sampled and analyzed in this study. Results showed that nucleotide variation was generally high within populations, and differentiation was very low among populations. Genetic diversity was slightly reduced in samples taken from the area with a recent strong reduction in population size. However, the allele frequency spectrum was skewed toward rare variants, and the association among variants of different sites was significantly more nonrandom within these samples compared with those from the area without such a population size reduction. The pattern was robustly supported by coalescent simulations.

PubMed ID: 19131702

The Journal of Heredity, 2005, Vol. 96, No. 1, pp. 1-10.

For modifying or submitting additional information



**Colleague Directory**

Search:    [Colleague | Laboratory]

Browse: - A - B - C - D - E - F - G - H - I - J - K - L - M - N - O - P - Q - R - S - T - U - V - W - X - Y - Z -

This database currently represents 1900 colleagues from 90 countries and 760 organizations

Colleague Name	Position	Organization	Update	Lab
<a href="#">add colleague</a> <a href="#">reset</a>				
HAFATI, Mohd	Research	INRA Domaine de la Grande Ferrade	05/08	IS
HOSE, DIBIRU	Postdoctoral Fellow	West Virginia University	02/09	WVU
Hu, Chun-Soo	Research Scientist	Forest Research Institute	11/04	FRRI
Huon, Campbell, Insook	Associate Professor	University of Georgia	02/09	UGA
Hakada, Ryogo	Research Scientist	National Forest Tree Breeding Center - Japan	01/07	HTBC
Hakahama, Hiroyuki	Breeder	Forestry and Forest Products Research Institute	2/05	FFRI
Hadi, Robert P.	Coordinator	Centre de Coordination du FORAFRI (Evreux)	01/01	CCFP
Hata, Tetsu	Research Scientist	National Forestry Japan		
Hauptke, Gella	Research Scientist	Federal Res. Forest Prod.		
Hachiyal, Subhash	Research Scientist	Forest Resear.		
Hawano, Carlos	Forest Scientist	CATIE Tropi. Research and		
Huydenov, Stasimir	Research Scientist	Forest Resear.		

**Representing ~2,000 colleagues from 90 countries and 760 organizations**

Registered individuals are linked with their publications, data sets

Directory allows users to add/modify their own record as well as register for services on the site/application server

Dr. Neale, David	
position	Principal Investigator
lab code	IFG1
organization	Institute of Forest Genetics
department	Department of Environmental Horticulture
url	<a href="http://dendrome.ucdavis.edu/NealeLab/index.html">http://dendrome.ucdavis.edu/NealeLab/index.html</a>
addr 1	One shields avenue
city	Davis
state	California
country	UNITED STATES
e mail	dneale@ucdavis.edu
phone	(530) 754-8431
fax	(530) 754-9366
genus	Picea
species	Pinus taeda

## CMap: Comparative Map Viewer

CMap was originally written for the Gramene project  
<http://www.gramene.org/>

Now part of the GMOD (Generic Model Organism Database) Project  
<http://www.gmod.org/>

Can be queried from three different access points:

- 1) **Comparative map viewer:** Choose a starting map and add comparative maps to the right or left
- 2) **Feature search:** Search for a particular feature by name, accession ID, species, or feature type
- 3) **Correspondence matrix:** View the number of correspondences among all maps and choose a pair to display in the viewer; continue by adding more comparative maps

### Feature Search

A feature in CMap is any element that can be placed on a map, either as a point or an interval.

#### Feature Search

Search On:

Search For:

\* Separate multiple names with commas or white space. Use "" or "" for wildcards. To find features with spaces in their names, surround the name in double quotes, e.g., "ABC DEF".

Restrict species:   
 Acacia mangium  
 Cryptomeria japonica  
 Eucalyptus globulus  
 Eucalyptus grandis

Restrict feature types:   
 AFLP  
 Centromere  
 Contig

**Feature: AFLP**  
 Accession ID: P01P  
 Color: darkblue  
 Shape: line

**Feature: Microsatellite**  
 Accession ID: P03a016  
 Color: green  
 Shape: line

**Feature: Centromere**  
 Accession ID: cen  
 Color: black  
 Shape: box

**Feature: Phenotype**  
 Accession ID: phen  
 Color: black  
 Shape: span

**Feature: Clone**  
 Accession ID: clone  
 Color: black  
 Shape: span

**Feature: Protein**  
 Accession ID: Protein  
 Color: darkred  
 Shape: line

**Feature: Contig**  
 Accession ID: contig  
 Color: black  
 Shape: span

**Feature: QTL**  
 Accession ID: QTL  
 Color: darkgreen  
 Shape: line

**Feature: EST**  
 Accession ID: EST  
 Color: darkcyan  
 Shape: line

**Feature: RAPD**  
 Accession ID: RAPD  
 Color: red  
 Shape: line

**Feature: ESTP**  
 Accession ID: ESTD  
 Color: white  
 Shape: line

**Feature: Read**  
 Accession ID: read  
 Color: black  
 Shape: direction-arrow

**Feature: Gene**  
 Accession ID: gene  
 Color: chocolate  
 Shape: line

**Feature: Back To Far Annot**  
 Accession ID: far\_annot  
 Color: black  
 Shape: span

**Feature: Interpolated Phenotype**  
 Accession ID: int\_phen  
 Color: green  
 Shape: diamond

**Feature: RFLP**  
 Accession ID: RFLP  
 Color: blue  
 Shape: line

**Feature: Isozyme**  
 Accession ID: isozyme  
 Color: black  
 Shape: line

**Feature: SCAR**  
 Accession ID: SCAR  
 Color: darkgrey  
 Shape: line

**Feature: Low LOD Marker**  
 Accession ID: lowlod  
 Color: brown  
 Shape: diamond

**Feature: SNP**  
 Accession ID: SNP  
 Color: brown  
 Shape: line

**Feature: Marker**  
 Accession ID: marker  
 Color: black  
 Shape: line

**Feature: SSR**  
 Accession ID: SSR  
 Color: orange  
 Shape: line

Items 1 to 3 of 3

Map Name	Related Maps	Related Maps per unit	Correspondences	Correspondences per unit	Start	Stop	SSR	SSR per unit	GBS	GBS per unit
Poptr_C05J_2000_01 (Eucalypt)	0	0	0	0	0.00	59.00	0	0	1	16
Poptr_C05J_2000_02 (Eucalypt)	0	0	0	0	0.00	63.00	3	36.18%	1	12
Poptr_C05J_2000_03 (Eucalypt)	0	0	0	0	0.00	202.00	2	19.42%	1	3.2
Poptr_C05J_2000_04 (Eucalypt)	0	0	0	0	0.00	44.00	1	22.73%	0	0
Poptr_C05J_2000_05 (Eucalypt)	0	0	0	0	0.00	37.00	0	0	0	0
Poptr_C05J_2000_06 (Eucalypt)	0	0	0	0	0.00	175.00	1	5.71%	0	0
Poptr_C05J_2000_07 (Eucalypt)	0	0	0	0	0.00	103.00	0	0	1	9.9
Poptr_C05J_2000_08 (Eucalypt)	0	0	0	0	0.00	170.00	0	0	1	5.8
Poptr_C05J_2000_09 (Eucalypt)	0	0	0	0	0.00	68.00	0	0	0	0

**TreeGenes**

Colleague Directory

Organizations

Species Database

Literature Database

EST Database

Comparative Mapping Database

Updates

## Comparative Mapping Database (CMap)

CMap Home Search Matrix Map Sets View Types Species Display Sidebars (On/Off)

### Map Set Info

Restrict by Species:  Restrict by Map Type:

Items 1 to 4 of 4.

**Pita\_UCD\_genetic\_0\_Eckert**

Abbreviated Name: Pita\_UCD\_genetic\_0  
 Accession ID: TG994  
 Species: Pinus taeda (Loblolly Pine)  
 Map Type: Genetic  
 Map Units: cM  
 Linkage Groups: 12  
 Published On: N/A

Paper: High-throughput genotyping and mapping of single nucleotide Mapping Pop: QTL and BASE pedigrees sensu Groover et al. (1994)

Linkage Groups	Map Viewer	Map Details	Matrix Viewer	Download Map Data
Pita_UCD_0_01	[ Map Viewer ]	[ Map Details ]	[ Matrix Viewer ]	[ Download Map Data ]
Pita_UCD_0_02	[ Map Viewer ]	[ Map Details ]	[ Matrix Viewer ]	[ Download Map Data ]
Pita_UCD_0_03	[ Map Viewer ]	[ Map Details ]	[ Matrix Viewer ]	[ Download Map Data ]
Pita_UCD_0_04	[ Map Viewer ]	[ Map Details ]	[ Matrix Viewer ]	[ Download Map Data ]
Pita_UCD_0_05	[ Map Viewer ]	[ Map Details ]	[ Matrix Viewer ]	[ Download Map Data ]
Pita_UCD_0_06	[ Map Viewer ]	[ Map Details ]	[ Matrix Viewer ]	[ Download Map Data ]
Pita_UCD_0_07	[ Map Viewer ]	[ Map Details ]	[ Matrix Viewer ]	[ Download Map Data ]
Pita_UCD_0_08	[ Map Viewer ]	[ Map Details ]	[ Matrix Viewer ]	[ Download Map Data ]
Pita_UCD_0_09	[ Map Viewer ]	[ Map Details ]	[ Matrix Viewer ]	[ Download Map Data ]
Pita_UCD_0_10	[ Map Viewer ]	[ Map Details ]	[ Matrix Viewer ]	[ Download Map Data ]
Pita_UCD_0_11	[ Map Viewer ]	[ Map Details ]	[ Matrix Viewer ]	[ Download Map Data ]
Pita_UCD_0_12	[ Map Viewer ]	[ Map Details ]	[ Matrix Viewer ]	[ Download Map Data ]

**Events**

2nd Annual Plant Breeding Workshop: An Initiative of the Plant Breeding Coordinating Committee (SCC-080)

Des Moines, Iowa  
June 16-18, 2008

93rd ESA Annual Meeting: Enhancing Ecological Thought by Linking Research and Education

The Midwest Airlines Center - Milwaukee, Wisconsin  
August 3-8, 2008

International Workshop: Population, Quantitative and Comparative Genomics of Adaptation in Forest Trees

Centro di Ecologia Alpha, Monte Bondone, Trento, Italy  
August 3-5, 2008

http://dendrome.ucdavis.edu **Dendrome**

### Feature Search

A feature in CMap is any element that can be placed on a map, either as a point or an interval.

**Feature Search**

Search On: Name

Search For:

\*Separate multiple names with commas or vertical bars. Use "" or "" for wildcards. To find features with spaces in the name, surround the name in double quotes, e.g., "abc def".

Restrict species: -All Species-

Acacia mangium  
Cryptomeria japonica  
Eucalyptus globulus  
Eucalyptus grandis

Restrict feature types: -All Feature Types-

AFLP  
Centomere  
Clone  
Contig

Submit    Reset

**Map Search**

Published: Both

Ref. Species: Black Cottonwood (Populus trichocarpa) Change Species

Ref. Set: Genetic: Potr\_Mansfield\_genetic\_2000\_Chen\_837

Name (optional):

Min # of Related Maps:

Submit

Reference  
Douglas Fir  
Paine\_UCD\_genetic\_2009

Paine\_UCD\_2009\_01  
C263

**Feature Types:**

— SFLP

— SNP

**Menu Symbol:**

- Map Set Info
- Map Details
- Matrix View
- Limit to One Map
- Delete Map Set
- Delete Map
- Flip Map
- Unflip Map
- New Map View

Items List of 8

Map Name	Related Map:	Related Map: per unit	Correspondences: per unit	Start	Stop	SSP	total	per unit	total
Paine_CSIU_2000_01 (epiclar)	0	0.00	0	0.00	19.00	0	0	0	1
Paine_CSIU_2000_02 (epiclar)	0	0.00	0	0.00	19.00	1	36,148	1	12
Paine_CSIU_2000_03 (epiclar)	0	0.00	0	0.00	192.00	2	19,426	1	19
Paine_CSIU_2000_04 (epiclar)	0	0.00	0	0.00	44.00	1	12,736	0	0
Paine_CSIU_2000_05 (epiclar)	0	0.00	0	0.00	17.00	0	0	0	0
Paine_CSIU_2000_06 (epiclar)	0	0.00	0	0.00	176.00	1	15,216	0	0
Paine_CSIU_2000_07 (epiclar)	0	0.00	0	0.00	101.00	0	0	1	9
Paine_CSIU_2000_08 (epiclar)	0	0.00	0	0.00	170.00	0	0	1	6
Paine_CSIU_2000_09 (epiclar)	0	0.00	0	0.00	18.00	0	0	0	0

## CMap Direct Linking - Features

[http://dendrome.ucdavis.edu/cmap/index.php?linkto=FeatureDetail&map\\_name=Pita\\_IFG\\_1994\\_01&feature\\_name=rflpPitaIFG\\_1154\\_b](http://dendrome.ucdavis.edu/cmap/index.php?linkto=FeatureDetail&map_name=Pita_IFG_1994_01&feature_name=rflpPitaIFG_1154_b)

**TreeGenes**

Colleague Directory

Colleagues

Organizations

Species Database

Forest Trees

Literature Database

Search Literature

Submit Publication

Submit Supplemental

EST Database

EST Database

EST Summary

Submit EST Data

Search ESTs

Resequencing Database

**Comparative Mapping Database (CMap)**

CMap Home   Search   Matrix   Map Sets   View Types   Species   Display Sidebars (On/Off)

**Feature "rflpPitaIFG\_1154\_b"**

Feature Name: rflpPitaIFG\_1154\_b

Aliases: N/A

Accession ID: 7911

Feature Type: AFLP

Map: Liberty Pine | Pita\_IFG\_genetic\_1994\_Devry\_273 | Pita\_IFG\_1994\_01

Start: 10.40 cM

Stop: 10.40 cM

**Correspondences**

No correspondences to show.

**Events**

Opportunities, challenges and limitations of genomic-based technologies in forest tree breeding and forest genetics

Freiburg, Germany  
October 7-9, 2009

Genomics of Forest and Ecosystems Health in the Fagaceae (Beech Family)

Research Triangle Park, North Carolina  
November 10-13, 2009

International Symposium on Biology of Rare and Endemic Plant Species

Mugla Turkey  
May 26-29, 2010



# CMap – Obtaining TG Accession Number

**Obtain TreeGenes Accession Number**

The following form is for the submission of genetic map files and will provide the submitter with a TreeGenes accession number that will reference the map in the TreeGenes database. This interface is intended for maps associated with a manuscript that will be submitted to a peer-reviewed journal.

If you would like to submit the mapping files for a manuscript that has already been published, please follow this link: [link](#)

Please enter preliminary information about your manuscript and upload your genetic map as one file per linkage group. Detailed instructions on the nomenclature and the formatting of the files can be found [here](#).

Following successful submission, you will receive a TreeGenes accession number on the final screen and via e-mail. Your [catalog](#) entry must be up to date in our colleague database to begin this submission process.

Please reference the TreeGenes database and the reference number received here in the manuscript intended for submission.

For more information about how to complete this form, please [click here](#).

**Personal Information**

First Name:   
 Author:   
 Author's Email:   
 Other Authors:

**Publication Information**

Title:   
 Journal:   
 Volume:   
 Issue:   
 Accession Title:

**Genetic Map**

File:   
 Organization:

**Add Supplemental Data**

For instructions on the acceptable formats for these fields, [click here](#).

**Organizational Info**

Author:   
 Paper:   
 Supplement Type:

**Genetic Map**

File:   
 Organization:

(optional) Add additional map files

Obtain TG Accession number!

**Pinus taeda genetic\_2009\_Erkort\_225**

Accession ID: T254

Map Type: Genetic  
 Map (Meta): cM  
 Linkage Group: 12  
 Published On: January, 2009

Pages: High-throughput genotyping and mapping of single nucleotide polymorphisms in loblolly pine (Pinus taeda L.)  
 Mapping Paper QTL and BACe pedigree maps Groover et al. (1984)

**Linkage Groups**

Pinx_UCD_2009_31	Map Viewer	Map Details	Matrix Viewer	Download Map Data
Pinx_UCD_2009_32	Map Viewer	Map Details	Matrix Viewer	Download Map Data
Pinx_UCD_2009_33	Map Viewer	Map Details	Matrix Viewer	Download Map Data
Pinx_UCD_2009_34	Map Viewer	Map Details	Matrix Viewer	Download Map Data
Pinx_UCD_2009_35	Map Viewer	Map Details	Matrix Viewer	Download Map Data

Add literature data and (first) map file

<http://dendrome.ucdavis.edu>

Dendrome

# GMOD Genome Browser

Currently houses 10 annotated *Pinus taeda* BAC sequences



Showing 142.4 kbp from BAC3, positions 1 to 142,351

**Instructions**  
[Bookmark this](#) | [Upload your own data](#) | [Hide banner](#) | [Share these tracks](#) | [Link to image](#) | [High-res image](#) | [Help](#) | [Close](#)

**Search**  
 Landmark or Region:  Search

**Reports & Analysis**  
 Annotate Restriction Sites | Configure... | Go

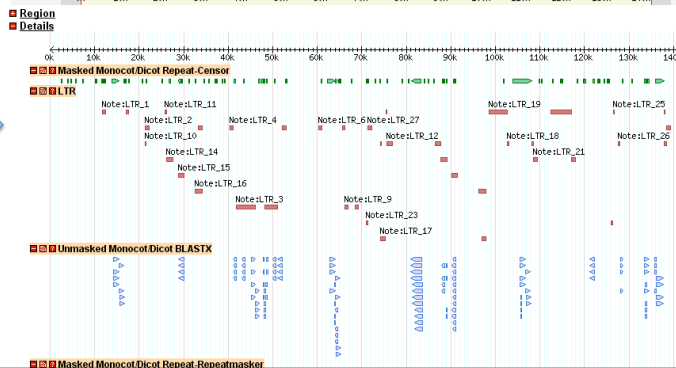
**Data Source**  
 Pinus taeda (loblolly pine) BAC3

ScrollZoom: <<< || Show 142.4 kbp >>> | Flip

**Overview**

Search and Select data source

Tracks can be reordered or hidden as necessary





# Literature Database and EST Database

represented >10 unique transcripts, did not show any significant homology to known sequences in the databases searched: these are expected to represent genes specific to *Cryptomeria* and, possibly, to related species.

Pubmed ID: 11052197

Keywords: CDNA, CONIFER, CRYPTOMERIA, CRYPTOMERIA JAPONICA, DNA, EST, FOREST TREE

**Supplemental Data:**

**Sequence Data**

→ All EST Data (30)

Plant Molecular Biology, 2000; Vol 43(4) : 451 - 457

**TreeGenes EST Database**

GenBank Accession ID | Link to NCBI | Link to TreeGenes EST Detail |

AU066349 [ NCBI   TG ]	AU066350 [ NCBI   TG ]	AU066351 [ NCBI   TG ]	AU066352 [ NCBI   TG ]	AU066353 [ NCBI   TG ]
AU066354 [ NCBI   TG ]	AU066355 [ NCBI   TG ]	AU066356 [ NCBI   TG ]	AU066357 [ NCBI   TG ]	AU066358 [ NCBI   TG ]
AU066360 [ NCBI   TG ]	AU066361 [ NCBI   TG ]	AU066362 [ NCBI   TG ]	AU066363 [ NCBI   TG ]	AU066364 [ NCBI   TG ]
AU066365 [ NCBI   TG ]	AU066366 [ NCBI   TG ]	AU066367 [ NCBI   TG ]	AU066368 [ NCBI   TG ]	AU066369 [ NCBI   TG ]
AU066370 [ NCBI   TG ]	AU066371 [ NCBI   TG ]	AU066372 [ NCBI   TG ]	AU066373 [ NCBI   TG ]	AU066374 [ NCBI   TG ]
AU066375 [ NCBI   TG ]	AU066376 [ NCBI   TG ]	AU066377 [ NCBI   TG ]	AU066378 [ NCBI   TG ]	AU066379 [ NCBI   TG ]

<http://dendrome.ucdavis.edu>



**Trace File Submission**

File Renaming

Submission of Project Details

- Contact Information
- Library Details
- Publication
- Protocols
- Reaction Conditions

**dbEST: database of "Expressed Sequence Tags"**

PubMed   Entrez   BLAST   OMIM   Taxonomy   Structure

dbEST Home

Access Links—

Entrez

BLAST Web

**Access to dbEST Data**

- EST sequences are included in the EST division of GenBank, available from NCBI by anonymous ftp and through Entrez.
- The nucleotide sequences may be searched using the BLAST electronic mail server. The TBLASTN program which takes an amino acid query sequence and compares it with six-frame translations of dbEST DNA sequences is particularly useful.
- EST sequences are also available as a flat file in the FASTA format by anonymous ftp in the /repository/dbEST directory at ftp.ncbi.nih.gov

**Sequence Processing**

FASTA Quality

Phred Base Calling

No Vectors

Mask (Cross\_Match)

> 200bp A

< 10bp

Trim PolyA tail

Clean Duplicates

Low Quality

Quality Clean

Bad Repeats

No Mito

Clean Mitochondrial

Results and MySQL Database Storage

FASTA Files

Mito Sequences

dbEST Files

High Quality Clean Sepp

Main Singlets

Main Contigs

Subgroups

New Subgroups

Cross-Species Database Storage

Enzyme Annotation

GO Annotation

Protein Domains

UTR Annotation

Repeat Annotation

Map Annotation

**Annotation**

Priority Blast

Parallel Blast

InterPro

With Hits

ECPProfiler

GOProfiler

Repeat Masker

BLAT Mapping

Public Submission

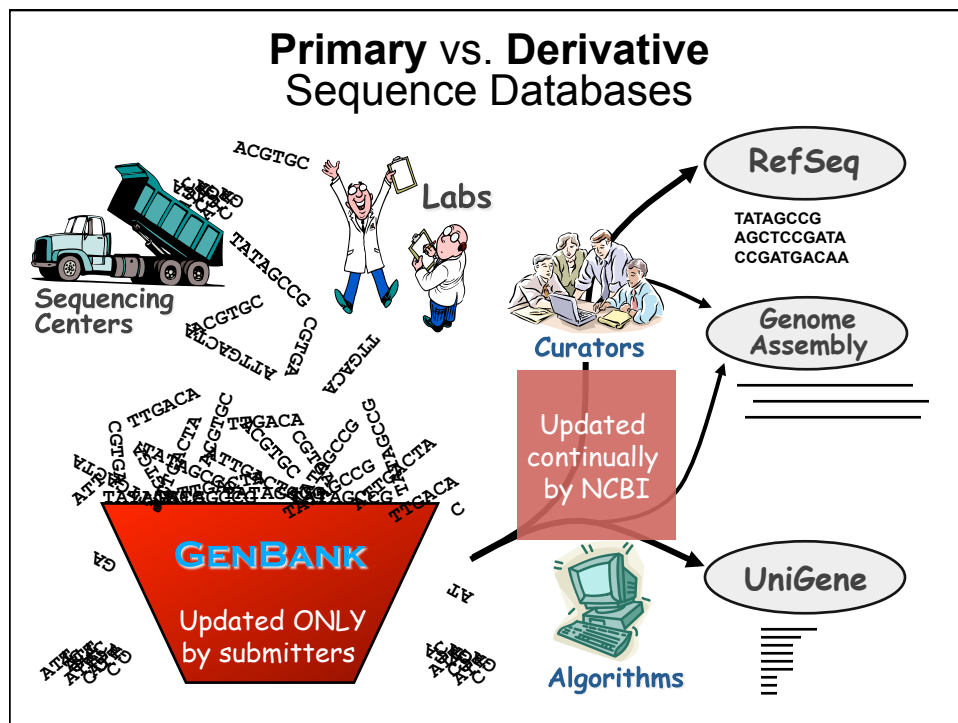
Statistics and Results Sent to User

Submission of Project Details to dbEST upon Approval

- Contact Information
- Library Details
- Publication
- Protocols
- Reaction Conditions

## ESTs: Annotation

- **BLAST against public protein and nucleotide databases**
  - Pairwise sequence alignment which indicates the region of the sequence which is similar.
- **Gene Ontology/Plant Ontology (structure & development)**
  - Standardized set of terms to describe gene products/ phenotypes
- **Protein domains**
  - The protein family databases (NCBI)
  - Searches for protein domains (InterPro).
- **Enzyme Annotation**
  - Enzyme Nomenclature formally defined for annotation
  - Links into databases such as BRENDA and KEGG
- **KEGG**
  - Large network of databases including chemical, gene and pathway
  - Provides annotation based on gene product interactions



# Plant Ontology

The **Plant Ontology Consortium (POC)** is a [collaboration](#) among plant genome databases and plant systematics to create, maintain and use controlled vocabularies to describe phenotypes and expression patterns of plant genes.

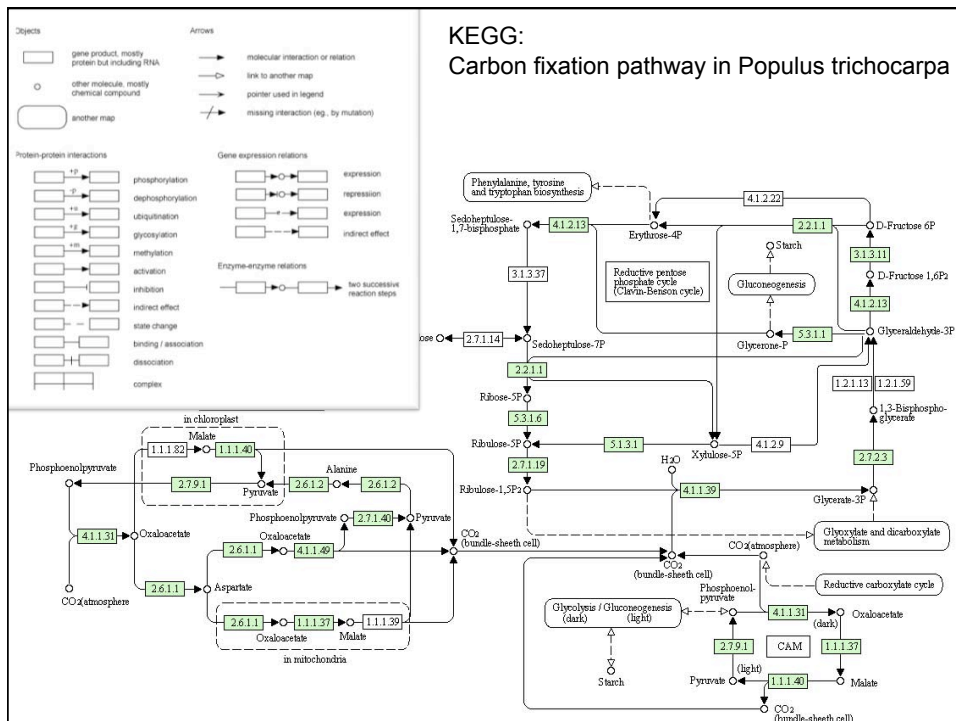
-Formed in response to the need for a set of uniform terms to describe plant structures and developmental stages.

The plant ontology comprises terms covering two biological domains:

**Plant Structure:** describes morphological/anatomical structures of plants including organs, tissues and cell types.

**Plant Development:** describes stages in the growth and development of an entire plant including development of individual organs, tissues and tissue systems.

<http://dendrome.ucdavis.edu>





## Sample Tracking Overview

### Objectives:

- 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
- Build a flexible system that can be applied to a variety of plant species (even animals!)

<http://dendrome.ucdavis.edu>

Dendrome 45

## Phenotypic Data Registration Process

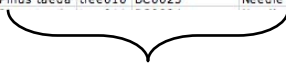
The image displays two screenshots of the Dendrome website's registration process. The left screenshot shows the 'TreeGenes User Account Request Form' with sections for 'Personal Info' and 'Contact Info'. The right screenshot shows the 'New User Registration' form with fields for 'Institution Name', 'Department', 'Name', 'Email', and 'Phone'. Below the screenshots is a barcode scanner and a barcode.

<http://dendrome.ucdavis.edu>


Dendrome

## Phenotypic Data Metadata


ImportData.csv										
	A	B	C	D	E	F	G	H	I	J
1	Species	Tree Id	Sample Barcode	S. Type	Tree Description	Age	GPS Coordinate	S. Description	BV-Cellulose	BV-Density
2	Pinus taeda	tree007	BC0001	Needle	Tall lobliolly tree	5	N30 17.477 W97 44.315	Sample 1	10	100
3	Pinus taeda	tree008	BC0002	Needle	Tall lobliolly tree	6	N30 17.477 W97 44.316	Sample 2	11.1	101
4	Pinus taeda	tree009	BC0003	Needle	Tall lobliolly tree	7	N30 17.477 W97 44.317	Sample 3	12.2	101
5	Pinus taeda	tree010	BC0004	Needle	Tall lobliolly tree	8	N30 17.477 W97 44.318	Sample 4	13.3	102
6	Pinus taeda	tree011	BC0005	Needle	NULL	NULL	N30 17.477 W97 44.319	NULL	NULL	NULL
7	Pinus taeda	tree012	BC0006	Needle	Tall lobliolly tree	10	N30 17.477 W97 44.320	Sample 6	15.5	400
8	Pinus taeda	tree013	BC0007	Needle	Tall lobliolly tree	11	N30 17.477 W97 44.321	Sample 7	16.6	999
9	Pinus taeda	tree014	BC0008	Needle	Tall lobliolly tree	12	N30 17.477 W97 44.322	Sample 8	17.7	998
10	Pinus taeda	tree015	BC0009	Needle	Tall lobliolly tree	13	N30 17.477 W97 44.323	Sample 9	55.5	100
11	Pinus taeda	tree015	BC0010	Needle	Tall lobliolly tree	13	N30 17.477 W97 44.324	Sample 10	99.9	858
12	Pinus taeda	tree007	BC0011	Needle	Tall lobliolly tree	9	N30 17.477 W97 44.325	Sample 5	14.4	103
13	Pinus taeda	tree008	BC0012	Needle	Tall lobliolly tree	10	N30 17.477 W97 44.326	Sample 6	15.5	103
14	Pinus taeda	tree009	BC0013	Needle	Tall lobliolly tree	11	N30 17.477 W97 44.327	Sample 7	16.6	104
15	Pinus taeda	tree010	BC0014	Needle	Tall lobliolly tree	12	N30 17.477 W97 44.328	Sample 8	17.7	104
16	Pinus taeda	tree011	BC0015	Needle	NULL	NULL	N30 17.477 W97 44.329	NULL	NULL	NULL
17	Pinus taeda	tree012	BC0016	Needle	Tall lobliolly tree	14	N30 17.477 W97 44.330	Sample 11	103.4	676
18	Pinus taeda	tree013	BC0017	Needle	Tall lobliolly tree	15	N30 17.477 W97 44.331	Sample 12	124.1	678
19	Pinus taeda	tree014	BC0018	Needle	Tall lobliolly tree	16	N30 17.477 W97 44.332	Sample 13	144.9	680
20	Pinus taeda	tree015	BC0019	Needle	Tall lobliolly tree	17	N30 17.477 W97 44.333	Sample 14	165.7	681
21	Pinus taeda	tree007	BC0020	Needle	Tall lobliolly tree	9	N30 17.477 W97 44.334	Sample 5	14.4	103
22	Pinus taeda	tree008	BC0021	Needle	Tall lobliolly tree	10	N30 17.477 W97 44.335	Sample 6	15.5	103
23	Pinus taeda	tree009	BC0022	Needle	Tall lobliolly tree	11	N30 17.477 W97 44.336	Sample 7	16.6	104
24	Pinus taeda	tree010	BC0023	Needle	Tall lobliolly tree	12	N30 17.477 W97 44.337	Sample 8	17.7	104




Primary Data



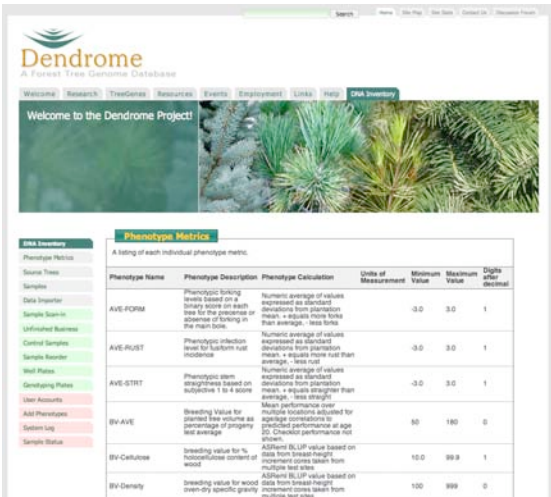
Metadata




Phenotype Data

<http://dendrome.ucdavis.edu>


## Phenotypic Data Phenotype Definitions



Phenotype Name	Phenotype Description	Phenotype Calculation	Units of Measurement	Minimum Value	Maximum Value	Digits after decimal
AVE-FORM	Phenotypic scoring levels based on a 5-point score on each tree for the presence or absence of forking in the main bole.	Numeric average of values increased as observed deviations from plantation mean + adjust more for less than average - less than		-0.0	3.0	1
AVE-RUST	Phenotypic infection level for uniform rust incidence	Numeric average of values increased as observed deviations from plantation mean + less rust		-0.0	3.0	1
AVE-STRT	Phenotypic stem straightness based on subjective 1 to 4 score	Numeric average of values increased as observed deviations from plantation mean + equals straighter than average - less straight		-0.0	3.0	1
BV-AVE	Breeding Value for normal tree volume percentage of progeny and average	Mean performance over multiple locations adjusted for sampling considerations		50	150	0
BV-Cellulose	Breeding value for % hemicellulose content of wood	Adjusted BLUP value based on data from breast height increment cores taken from multiple trees		10.0	99.9	1
BV-Density	Breeding value for wood oven-dry specific gravity	Adjusted BLUP value based on data from breast height increment cores taken from multiple trees		100	999	0

<http://dendrome.ucdavis.edu>






# Phenotypic Data

## Intuitive Data Filtering

A Forest Tree Genome Database

Welcome to the Dendrome Project!

**DNA Inventory**

**Samples**

A listing of each individual sample submitted for sequencing.

Date	Tree ID	Bar Code	Sample Type	Phenotype	Quantity	Status
2009.2.2	tree007	BC11121	Needle	AF	100	
				BV-Celulose	-99.9	
				BV-Density	-99.9	
				BV-Forn	10.0	
				BV-H3	10.0	
				BV-H8	-99.9	
				BV-Lignin	-99.9	
				BV-Fluor	-99.9	
				BV-Stn	0.000	
				BV-Taper-8	-99.9	
				BV-Vol-3	-99.9	
				BV-Vol-6	-99.9	
2009.2.2	tree008	BC11122	Needle	BV-Forn	0.0	
				BV-Celulose	11.1	
				BV-Density	101	
				BV-Forn	0.0	
				BV-H3	0.0	
				BV-H8	0.0	
				BV-Lignin	10.0	
				BV-Fluor	0.0	
				BV-Stn	0.000	
				BV-Taper-8	0.000	
				BV-Vol-3	0.0	
2009.2.2	tree009	BC11123	Needle	BV-Lignin	11.0	
				BV-Celulose	12.7	
				BV-Density	101	
				BV-Forn	-99.8	
				BV-H3	-99.8	
				BV-H8	-99.8	
				BV-Fluor	-99.8	
				BV-Stn	-50.0	
				BV-Taper-8	0.000	
				BV-Vol-3	-25.0	
				BV-Vol-6	-88.3	

<http://dendrome.ucdavis.edu>

# Sample Prep and DNA Isolation

## Well Plate Map with Concentrations

**DNA Inventory**

**96-Well Plate**

DWPP/PCR plate mapping and concentrations data.

DWPP/PCR Plate Map:

	1	2	3	4	5	6	7	8	9	10	11	12
A	1.0000 ng	0.5000 ng	0.2500 ng	0.1250 ng	0.0625 ng	0.0312 ng	0.0156 ng	0.0078 ng	0.0039 ng	0.0019 ng	0.0009 ng	0.0005 ng
B	1.0000 ng	0.5000 ng	0.2500 ng	0.1250 ng	0.0625 ng	0.0312 ng	0.0156 ng	0.0078 ng	0.0039 ng	0.0019 ng	0.0009 ng	0.0005 ng
C	1.0000 ng	0.5000 ng	0.2500 ng	0.1250 ng	0.0625 ng	0.0312 ng	0.0156 ng	0.0078 ng	0.0039 ng	0.0019 ng	0.0009 ng	0.0005 ng
D	1.0000 ng	0.5000 ng	0.2500 ng	0.1250 ng	0.0625 ng	0.0312 ng	0.0156 ng	0.0078 ng	0.0039 ng	0.0019 ng	0.0009 ng	0.0005 ng
E	1.0000 ng	0.5000 ng	0.2500 ng	0.1250 ng	0.0625 ng	0.0312 ng	0.0156 ng	0.0078 ng	0.0039 ng	0.0019 ng	0.0009 ng	0.0005 ng
F	1.0000 ng	0.5000 ng	0.2500 ng	0.1250 ng	0.0625 ng	0.0312 ng	0.0156 ng	0.0078 ng	0.0039 ng	0.0019 ng	0.0009 ng	0.0005 ng
G	1.0000 ng	0.5000 ng	0.2500 ng	0.1250 ng	0.0625 ng	0.0312 ng	0.0156 ng	0.0078 ng	0.0039 ng	0.0019 ng	0.0009 ng	0.0005 ng
H	1.0000 ng	0.5000 ng	0.2500 ng	0.1250 ng	0.0625 ng	0.0312 ng	0.0156 ng	0.0078 ng	0.0039 ng	0.0019 ng	0.0009 ng	0.0005 ng

DNA Mass Threshold: 1000 ng  
Volume of sample per well: 200 ul

Concentration Standards Regression Plot

$y = 0.0001x + 0.0000000000$   
 $R^2 = 0.9999999999$

<http://dendrome.ucdavis.edu>

## Sample Prep and DNA Isolation Genotyping Plate Creation

The screenshot shows the 'Add Genotyping Plate' interface in the Dendrome web application. The top navigation bar includes 'Welcome', 'Research', 'TreeGenes', 'Resources', 'Events', 'Employment', 'Links', 'Help', and 'DNA Inventory'. A sidebar on the left lists various functions like 'Phenotype Metrics', 'Source Trees', 'Samples', 'Data Importer', 'Sample Scan-in', 'Unfinished Business', 'Control Samples', 'Sample Recorder', 'Well Plates', 'Genotyping Plates', 'User Accounts', 'Add Phenotypes', 'System Log', and 'Sample Status'. The main form area is titled 'Add Genotyping Plate' and contains the following fields:

- Date: 6/16/2009
- Genotyping Plate Type: Illumina 96-Well Plate
- Genotyping Plate Barcode: W00099028
- Intra-Plate Control Samples (Optional): Control Sample: --NONE--, DNA Mass: [ ] ng
- Inter-Plate Control Samples (Optional): Control Sample: CTRL\_01, DNA Mass: [1100] ng
- Well Plate Exclusions (Well plates that should NOT be considered as sources of samples):  CTGN06

A 'Continue' button is located at the bottom of the form.

<http://dendrome.ucdavis.edu>

## Exported Genotype Data

The diagram illustrates the process of exporting genotype data. It starts with a stack of yellow folders on the left, representing the source data. An arrow points to a screenshot of a data analysis software window (likely GenAlEx) in the center, which displays a table of genotype data and a scatter plot. Another arrow points to a stack of white document pages on the right, representing the exported data files.

<http://dendrome.ucdavis.edu>

## Genotyping Data Download

The screenshot shows the Dendrome web interface. On the left is a navigation menu with items like 'DNA Inventory', 'Phenotype Metrics', 'Source Trees', 'Samples', 'Genotyping Data', 'Data Importer', 'Sample Scan-in', 'Unfinished Business', 'Control Samples', 'Sample Reorder', 'Well Plates', 'Genotyping Plates', 'User Accounts', 'Add Phenotypes', 'System Log', 'Sample Status', and 'Data Imports Summary'. The 'Genotyping Data' section is active, showing a 'Download' button and a 'Total Processing Time: 1 Min 23 Sec'. A file named 'GenotypingData\_20091001152846.csv' is being opened in Firefox, with the 'Save to Disk' option selected.

<http://dendrome.ucdavis.edu>

Dendrome

## PineSAP (Sequence Alignment and SNP Identification) & DNASam (DNA Sequence Analysis and Manipulation)

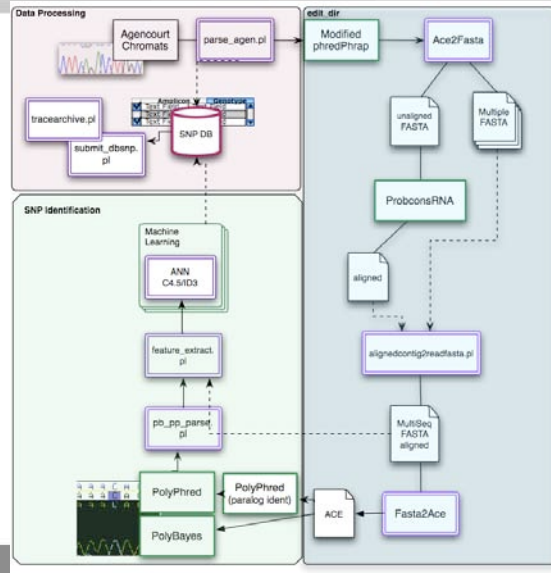
Wegrzyn J.L., Lee J.M., Liechty J., Neale D.B. (2009). PineSAP - Pine alignment and SNP Identification Pipeline. *Bioinformatics*

Eckert A.J., Tearse B.R., Pande B., Neale D.B. (2009). DnaSAM: Software to perform neutrality testing for large data sets with complex null models. *Molecular Ecology Resources*.

<http://dendrome.ucdavis.edu>

Dendrome

## PineSAP – Sequence Alignment and SNP Identification



Re-Sequencing data from Agencourt:

**Initial Processing**

**Base Calling**

**Sequence Alignment**

**SNP Identification**

**Machine Learning**

**Data Storage & Release**

<http://dendrome.ucdavis.edu>



## Alignment and SNP Identification Evaluation Criteria



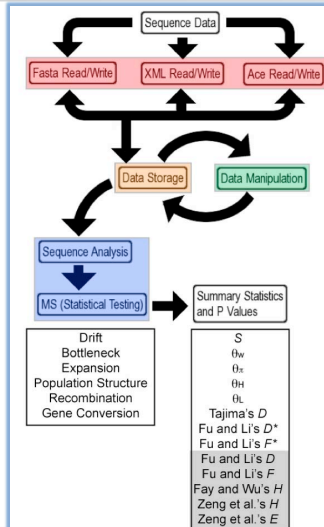
Evaluation	J48	Polyphred	Polybayes
Accuracy	93.6	76.25	78.02
Sensitivity	88.21	83.22	86.54
Specificity	98.73	N/A	N/A

Accuracy = (TP + TN)/total  
 Sensitivity = TP/(TP + FN)  
 Specificity = TN/(FP + TN)

Evaluation	J48	Polyphred	Polybayes
Accuracy	94.6	79.35	80.24
Sensitivity	90.54	85.01	88.14
Specificity	97.23	N/A	N/A



## DnaSAM - DNA Sequence Analysis and Manipulation



- Addresses the challenges of data manipulation, summary statistic estimation and statistical hypothesis testing for large-scale resequencing projects.
- Uses site frequency spectrum to quantify patterns of DNA polymorphisms observed within populations
- Can be used to understand the process of demography, adaptation, and speciation
- **Capable of performing a large number of standard and newly designs tests of neutrality for multiple sequence alignments of resequenced gene loci**

<http://dendrome.ucdavis.edu>

Dendrome

DiversiTree

<http://dendrome.ucdavis.edu>

Dendrome

## 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)

<http://dendrome.ucdavis.edu>

Dendrome

## Search Components

### Search Pane

Searches

+ EST (GenBank)

+ Contig

+ Sequence Assembly

Sequence ID: 0\_815  
(Ex: 0\_8156\_02)

Mark all species of interest:

- Picea abies (Paab)
- Pinus elliotii (Piel)
- Pinus lambertiana (Pila)
- Pinus pinaster (Pipn)
- Pinus radiata (Pira)
- Pinus sylvestris (Pisy)
- Pinus taeda (Pita)
- Pseudotsuga menziesii (Psme)

Results: 4

+ Amplicon

+ Primer

### Results Window

ID	Sequence ID	Species	Project	Has SNPs
4560	0_8155_01	Pita - Pinus taeda (Loblolly Pine)	PineSAP-ADEPT2	Yes
582	0_8156_01	Pita - Pinus taeda (Loblolly Pine)	PineSAP-ADEPT2	Yes
7270	0_8156_02	Pita - Pinus taeda (Loblolly Pine)	PineSAP-ADEPT2	Yes
3581	0_815_01	Pita - Pinus taeda (Loblolly Pine)	PineSAP-ADEPT2	Yes

### Detail Window

Assembly Name: 0\_815\_01 20060927 Pita  
Assembly Length: 233  
Assembly Protocol ID: 9

Notes

Contig/Species: Pinus / Pinus taeda  
SNP SNP ID: sig-Pita-0\_815\_01-47  
Parental: PineSAP-ADEPT2  
Candidate: AC  
Length: 5  
ACE File: [Download](#)  
FASTA File: [View](#)

Chromatogram Archive: [Download](#)

Sequencing Chromatogram	Tree Sample	View Chromatogram
Sequencing Chromatogram 0_8155_01-1667_Pi.F	Tree Sample: 17916	<a href="#">View Chromatogram</a>
Sequencing Chromatogram 0_8155_01-1667_Pi.F	Tree Sample: 17916	<a href="#">View Chromatogram</a>
Sequencing Chromatogram 0_8155_01-1668_Pi.F	Tree Sample: 22001	<a href="#">View Chromatogram</a>
Sequencing Chromatogram 0_8155_01-1668_Pi.F	Tree Sample: 22001	<a href="#">View Chromatogram</a>
Sequencing Chromatogram 0_8155_01-1669_Pi.F	Tree Sample: 21011	<a href="#">View Chromatogram</a>
Sequencing Chromatogram 0_8155_01-1669_Pi.F	Tree Sample: 21011	<a href="#">View Chromatogram</a>
Sequencing Chromatogram 0_8155_01-1670_Pi.F	Tree Sample: 21011	<a href="#">View Chromatogram</a>
Sequencing Chromatogram 0_8155_01-1670_Pi.F	Tree Sample: 21011	<a href="#">View Chromatogram</a>
Sequencing Chromatogram 0_8155_01-1671_Pi.F	Tree Sample: 23009	<a href="#">View Chromatogram</a>
Sequencing Chromatogram 0_8155_01-1671_Pi.F	Tree Sample: 23009	<a href="#">View Chromatogram</a>
Sequencing Chromatogram 0_8155_01-1672_Pi.F	Tree Sample: 19002	<a href="#">View Chromatogram</a>
Sequencing Chromatogram 0_8155_01-1672_Pi.F	Tree Sample: 19002	<a href="#">View Chromatogram</a>

### Data Window

ACE File: AS 1\_36

```

CO 0_8155_01 233 36 2 U
CCCCCACTGACGCAAAATACACTACACAGCCAAATCCGCCCTTCMATA
CAATCAAAAGAGCAATACAGTACCTACAAATGGCCGATGGGATTT
TAGTACTATTTGAGCCGCAAGTTAATGTCGACCTCTGATTTTAA
ATTCTTACCTTATACCTCTGACATTAAGCCAGCCGCAATACAC
TCATAATACGCAATGCAATACCTGCTTCTG
  
```

<http://dendrome.ucdavis.edu>

Dendrome

## Search Panes

**+ EST (GenBank)**

EST Name:

GenBank Accession:

GenBank Cl:

Select the species of interest:

Results: 129422

**+ Contig**

Contig Name:

BLAST GenBank Accession:

BLAST GenBank Cl:

BLAST Hit:

CO Term:

InterPro Term:

Enzyme:

Mark all species of interest:

Results: 84707

**+ Amplicon**

Amplicon Name:

BLAST Hit:

CO Term:

Enzyme:

Results: 21416

**+ Sequence Assembly**

Sequence ID:

Mark all species of interest:

Results: 12362

**+ SNP**


Sequence ID:

EST Contig:

Mark all species of interest:

Results: 66328

<http://dendrome.ucdavis.edu>



**Searches**

**+ EST (GenBank)**

**+ Contig**

Contig Name:

BLAST GenBank Accession:

BLAST GenBank Cl:

BLAST Hit:

CO Term:

InterPro Term:

Enzyme:

Mark all species of interest:

Results: 8

**+ Tree Samples**

**+ DNA Extractions**

**+ Seq Chromats**

**+ Sequence Assembly**


**+ SNP**

**+ Amplicon**

**+ Primer**

**+ Sequencing Run**

**+ Phenotype**





The screenshot shows the DiversiTree search interface. On the left, there are search filters for Contig Name, BLAST GenBank Accession, and BLAST Hit. The BLAST Hit is set to 'Glycogen (E. coli)'. Below these are checkboxes for various species like Pinus sibirica, Pinus lambertiana, etc. On the right, a table displays BLAST results with columns for Contig Name, Species, Function, and two 'Yes/No' columns. The table lists several entries for Pinus taeda and Pinus sibirica related to glycogen synthase 3.

Contig Name	Species	Function	Yes	No
11011_0_11113	Pinus taeda (Siberia) Pinus (glycogen synthase kinase 3)		Yes	No
11099_0_4981	Pinus taeda (Siberia) Pinus (glycogen synthase kinase 3)		No	No
12314_0_1883	Pinus taeda (Siberia) Pinus (glycogen synthase kinase 3)		No	No
14513_0_8126	Pinus taeda (Siberia) Pinus (glycogen synthase kinase 3)		No	No
21987_2_10567	Pinus taeda (Siberia) Pinus (glycogen synthase kinase 3)		No	No
20227_2_4796	Pinus taeda (Siberia) Pinus (glycogen synthase kinase 3)		No	No
28151_2_8178	Pinus taeda (Siberia) Pinus (glycogen synthase kinase 3)		No	No

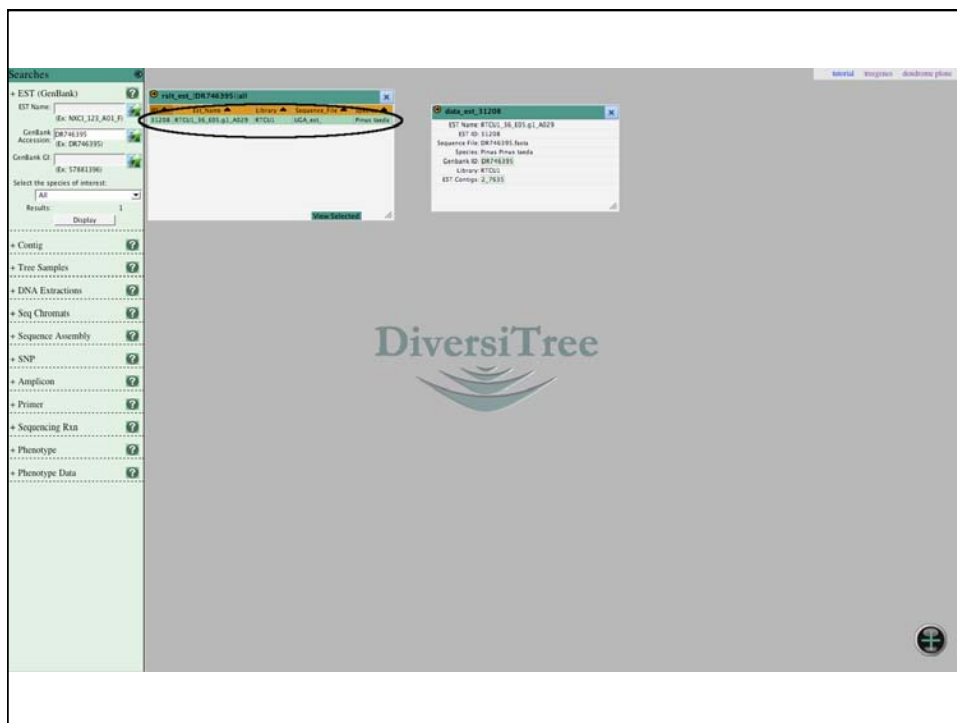
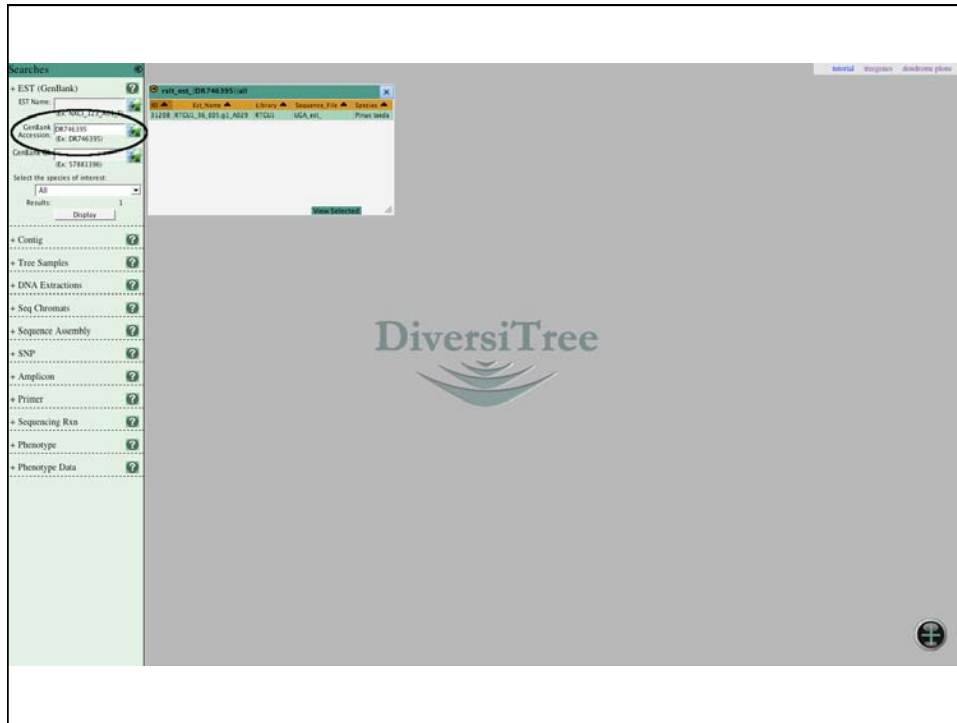
This screenshot shows the detailed BLAST results for contig 11113. A pop-up window titled 'data\_rowing\_1410' displays the following information:

- Contig Name: 0\_11113
- ID: 1410
- Sequence File: 0\_11113.fasta
- Gene/Source: Pinus / Pinus taeda
- Gene: RTCSL\_E2\_411\_3\_1\_A028
- Forward Primer: 0\_11113\_81\_4
- Reverse Primer: 0\_11113\_81\_4
- Amplifier: 0\_11113\_01
- Seq Assembly: 0\_11113\_01\_20061004.fasta
- BLAST Hit: glycogen synthase kinase 3
- BLAST GenBank ID: 224119492
- BLAST GenBank Accession: AC058452

Below this, it lists various Gene Ontology (GO) terms such as 'Function: GO Term: protein kinase activity', 'Function: GO Term: ATP binding', and 'Function: GO Term: protein serine/threonine kinase activity'. The sequence itself is displayed at the bottom of the window.

The screenshot displays a bioinformatics search interface. On the left is a sidebar with search filters for EST (GenBank), Contig, BLAST GenBank Accession, BLAST GenBank Q, BLAST Hit, GO Term, InterPro Term, Enzyme, and Mark all species of interest. The main area shows search results for 'data\_sanger\_1410'. A table lists search results with columns for Contig Name, Species, BLAST hit, and two checkboxes. Below the table, a detailed view for 'data\_sanger\_1410' shows sequence information, BLAST GenBank Accession (AC026167), and a list of GO terms. A second window shows 'data\_sangerassembly\_4030' with a list of sequencing events, including Chromosome, Time Sample, and View Chromatogram links.

This screenshot shows the same bioinformatics search interface as above, but with an additional window open. The 'data\_chromatogram\_ADEPT2\_0\_11113\_01-0051\_P1-F' window displays a chromatogram plot with a sequence alignment overlaid. The plot shows peaks for each base (A, C, G, T) and the corresponding sequence. The 'data\_sanger\_1410' and 'data\_sangerassembly\_4030' windows are also visible, showing the same search results and sequencing event lists as in the first screenshot.





The screenshot displays the Dendrome database search interface. On the left, a sidebar contains search filters for EST (GenBank), Contig, Tree Samples, DNA Extractions, Seq Chromatograms, Sequence Assembly, SNP, Amplicon, Primer, Sequencing Run, Phenotype, and Phenotype Data. The main area shows search results for 'data\_est\_287483991.tst' and 'data\_est\_31208'. A detailed view for 'data\_est\_31208' is shown, including its EST Name, ID, Sequence File, Species, Contig ID, and EST Contig. Below this, three 'data\_est\_31208' entries are listed with their respective physical positions, lengths, and flanking sequences. On the right, a 'data\_contig\_29122' entry is shown with its Contig Name, ID, Sequence File, Genus/Species, ESTs, Forward/Reverse Primers, Amplicon, Seq Assembly, and BLAST results. The BLAST results include the hit name, accession number, and alignment details.

## Bulk Retrieval Window

The screenshot shows the 'Bulk Retrieval Window' in the Dendrome database interface. The background displays the search sidebar and the 'DiversiTree' logo. Overlaid on the right is the 'EST Config List Uploader' dialog box. This dialog box has a 'Upload File' section with a 'Browse...' button. Below this, there are several sections for selecting data columns to include in the final output file: 'Basic Data', 'Functional Annotation Data', 'EST Contig Data', and 'Species Filter'. Each section contains a list of checkboxes for various data fields. The 'Basic Data' section includes Contig Name, Species, and Sequence Length. The 'Functional Annotation Data' section includes BLAST Description, Similarity Score, E-Value, SignalP, TrnScan, GO Term (Biological Process), InterPro, and GO Term (Molecular Function). The 'EST Contig Data' section includes Sequence (FASTA Format), Reannotated (Yes/No), SNPs (Yes/No), Primer IDs, and Amplicon IDs. The 'Species Filter' section includes a dropdown menu to select the species to include. A 'Done' button is located at the bottom of the dialog box.

<http://dendrome.ucdavis.edu>

Dendrome

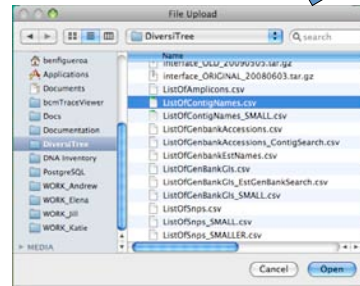
## Bulk Retrieval Window Components

### Bulk Retrieval Window



### File Selector

**Upload File:**  
 Select an Excel CSV file for upload. This file should contain a single column list of Contig Names.

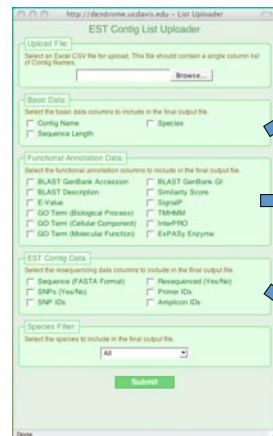


<http://dendrome.ucdavis.edu>



## Bulk Retrieval Window Components

### Bulk Retrieval Window



### Data & Annotation Selection Fields

**Basic Data:**  
 Select the basic data columns to include in the final output file.

Contig Name       Species  
 Sequence Length

**Functional Annotation Data:**  
 Select the functional annotation columns to include in the final output file.

BLAST GenBank Accession       BLAST GenBank GI  
 BLAST Description       Similarity Score  
 E-Value       SignalP  
 GO Term (Biological Process)       TMHMM  
 GO Term (Cellular Component)       InterPRO  
 GO Term (Molecular Function)       ExpASY Enzyme

**EST Contig Data:**  
 Select the resequencing data columns to include in the final output file.

Sequence (FASTA Format)       Resequenced (Yes/No)  
 SNPs (Yes/No)       Primer IDs  
 SNP IDs       Amplicon IDs

<http://dendrome.ucdavis.edu>



## Comments, Concerns, Help!



**dendrome.ucdavis.edu**  
Dendrome Help Forum

---


[Login](#) [Register](#)  
[View unanswered posts](#) | [View active topics](#)  
[Board index](#)

---

Forum

---

**Discussions**

-  **Help, it doesn't work**  
 For things that don't seem to work  
 Moderator: **btearse**
-  **Suggestions for TreeGenes**  
 Thoughts and discussions about what we could make better.
-  **General Usage Questions**  
 Questions about how things work or why they function the way they do.
-  **Plant Genomics**  
 Discussions of Plant Genomics Developments, Strategies, and Methods
-  **Bioinformatics**  
 Talk about old and new methods or technologies
-  **Open Source Development**

---

**Forum**

-  **Strobosphere**

Delete all board cookies | The team



## Future Directions: Dendrome Project and TreeGenes Database

### Objectives:

- Expansion of DiversiTree
  - Full integration of phenotypic data
  - Standardization of phenotypic data
- Continued expansion to incorporate available data from all forest trees
- Completion of a Genetic Stock Center
- More Sequence
  - Integration of sequence from high-throughput technologies
  - Better integration with genome databases



**\* Our Team**

**PROJECT DIRECTOR**  
David Neale – [dbneale@ucdavis.edu](mailto:dbneale@ucdavis.edu)

**LEAD BIOINFORMATICIAN**  
Jill Wegrzyn – [jlwegrzyn@ucdavis.edu](mailto:jlwegrzyn@ucdavis.edu)

**DEVELOPERS**  
Ben Figueroa  
Minyoung Choi  
John Liechty

<http://dendrome.ucdavis.edu> 