# TreeGenes & Tripal



treegenesdb.org

Emily Grau Department of Ecology & Evolutionary Biology University of Connecticut, Storrs CT

# TreeGenes Database: History

- Began as the Dendrome project (USDA funded initiative) in 1993 to hold forest tree genetic maps and associated markers
- One of the 1<sup>st</sup> UDSA funded databases on the internet
- Schema has been changing and evolving ever since



# TreeGenes Database: History

- Began to hold forest tree genetic maps and associated markers
- Expanded to other data types
  - Sequence
    - Resequencing, Large-Scale Genotyping, Transcriptomics/Expression
    - Full Genome Sequences
  - Analysis and Visualization Tools
    - Ability for users to mine the data
  - Resources for the user community
    - Literature, Colleagues



# TreeGenes Database: Users

treegenesdb.org

## 2,086 users from 862 organizations in 94

countries

10,000 -

Unique Web Visitors to TreeGenes Database per month, January-December

# TreeGenes Database: Species

- 1,774 species from 101 genera
  - At least one genetic artifact from each species
  - Conifers but is also inclusive of all forest trees
- Full genome sequence: 13 species
- Transcriptome/Expression resources: 3,920,817 sequences from 263 species
- 106 genetic maps from 35 species

treegenesdb.org

#### **Automated scripts**

- Primary databases such as NCBI
- Appropriate data should be submitted to primary databases first

#### **User submissions**

• For data and metadata not captured well by primary databases

treegenesdb.org

## **Automated scripts**

- NCBI
  - Transcripts, Protein, Unigene
     Databases
  - Linked to literature records, etc.
  - Incorporated into visualization tools
- Literature
  - Web of Science, PubMed

NCBI - <PubmedArticleSet> - <PubmedArticle> - <MedlineCitation Owner="NLM" Status="PubMed-not-MEDLINE"> <PMID Version="1">25202539</PMID> - <DateCreated> <Year>2014</Year> <Month>09</Month> <Dav>09</Dav> </DateCreated> - <DateCompleted> <Year>2014</Year> <Month>09</Month> <Dav>09</Dav> </DateCompleted> - <DateRevised> <Year>2014</Year> <Month>09</Month> <Day>12</Day> </DateRevised> - <Article PubModel="Electronic-eCollection"> - < Iournal> <ISSN IssnType="Electronic">2168-0450</ISSN> - <JournalIssue CitedMedium="Print"> <Volume>1</Volume> <Issue>4</Issue> - <PubDate> <Year>2013</Year> <Month>Apr</Month> </PubDate> </JournalIssue> <Title>Applications in plant sciences</Title> <ISOAbbreviation>Appl Plant Sci</ISOAbbreviation> </Journal> - <ArticleTitle> Application of proteomics to the study of pollination drops </ArticleTitle> - <Pagination> <MedlinePgn/> </Pagination> <ELocationID EIdType="doi" ValidYN="Y">10.3732/app <ELocationID EIdType="pii" ValidYN="Y">apps.1300003 – <Abstract> <AbstractText Label="UNLABELLED">•</AbstractTe - <AbstractText Label="PREMISE OF THE STUDY" NIR Pollination drops are a formative component in gymnos sexual reproduction. • </AbstractText>

 - <AbstractText Label="METHODS" NlmCategory="ME Pollination drops were sampled from eight gymnosperm invited benchmarked.

treegenesdb.org

### **User submissions**

- Internal projects or collaborations (day one)
- Submissions of data post-analysis at publication time

treegenesdb.org

#### User submissions

TGAD Accession Request Form			
This submission process will allow you to upload gene TreeGenes database. Results of the association tests collectively available using a TGAD accession. Result	will be accepted as raw P-	values and all files will be	
collectively available using a TGAD accession. Result publication of the associated manuscript.	s can be released immedia		
Species		TGAD Accession Request Form	
Please fill in the fields below for data pertaining to Ple If your study includes multiple species, each set of sp	ecies data will be collecte	Species Please fil in the fields below for data pertaining to Please teeds samples	anhe .
Genetic Data		If your study includes multiple species, each set of species data will be o	solected one at a time.
Does your study include Genetic data?		GPS Data	
Yes	-	Select the statement that describes how GPS coordinate data was used GPS coordinates were not collected in this study.	*
Phenotype Data		Phenotype Data	-
Does your study include Phenotype?		Select the statement that describes how Phenotype data was used in yo	
Yes	<u>*</u>	Defined Phenotype measurements were performed on each genotyped	3
Environmental Data Does your study include Environmental data?		Environmental Data Select the statement that describes how Environmental data was used in	1 your study:
Yes		Environmental measurements were performed on each genotyped indiv	a
-		Association Results Data Select the statement that best describes your Association results data:	
Co	ntinue	Single Marker	-
L		Continue	
		Continue	
		Continue	
		Cortrue	]
COMMUTIVICIAN Advances Tables		Control	
The Address has a built of sector of sector	and they are the second to	Control	
This full-bency form is for the outprocess of genetic the outprober with a first-bench accession running is first first-bench accession. This interface is each with a menuacing that will be submitted to a perior with a menuacing that will be submitted to a perior	nagi filas and wit provide that wit reflector de mas dati for mags annocided rviceed journal.	Contrue	
This following to in to for the outprotour of genetic the outprotes with a financial procession municipal of the financial protocols. This section is repre- went a non-constraint will be abilitated to a poten- tic and an additional financial for a section of the section of the financial financial financial financial fields for additional financial fields.	read from and wit provide that will refer to not the mass and for mass associated overed journal. answingt that has arready	Contrue	
This full-bency form is for the outprocess of genetic the outprober with a first-bench accession running is first first-bench accession. This interface is each with a menuacing that will be submitted to a perior with a menuacing that will be submitted to a perior	read from and wit provide that will refer to not the mass and for mass associated overed journal. answingt that has arready	Contra	
The science space is to the undersource of genues, the science with a first first sector sector sector is the target of the science of the science of the science of the science science of the science of the science of the science of the science of the science first science of the science of the science of the science science of the science of the science of the science science of the science of the science of the science science of the science of the science of the science science of the science of the science of the science science of the science of the science of the science science of the science of the science of the science science of the science of the science of the science of the science science of the scienc	hap has and wit provide but wit reference the way do for many service and indevelopment. Another that has arready incomed and wipcall your intervections on the many closes.	Cona	
The spokers parks for the admission of provid- tion and the site of the sectors without the sectors without the sectors without the sectors without the spokers of the sector sectors with the spokers of the sector sectors with sectors and the sector sector sectors with sectors and the sector sectors with sectors and the sectors with sector	map has seed with provide but with the provide map to be any any second second weight of any second second weight of the second second managed and second second managed and seco	Cona	
In the operation is a work of the design of the second sec	nachina ord wit provide beginn information the was able for most processed able for most processed another that has arrested in proceed point in proceed point in proceed point in the process and the set is the process.	Contra	
In the stear of prior A, by the submet of a start of the stear of the start of the	nachina ord wit provide beginn information the was able for most processed able for most processed another that has arrested in proceed point in proceed point in proceed point in the process and the set is the process.	Color	
A subject of a sky by advances of gauge to be a strength of a sky by advances of gauge to be a strength of a sky by advances of the sky by a sky by	nachina ord wit provide beginn information the was able for most processed able for most processed another that has arrested in proceed point in proceed point in proceed point in the process and the set is the process.	Contra.	
A second	man free ord wit provide book in information the way obtain information and the provided sources in incorpt that has areador in incorpt that has areador in incorpt and uptical pro- trained and information information and the source information and the information and the information of the sources information of the sources information of the sources information and the information of the sources information and the information and the inform	Lotha	
A second	manifestor el el production de la versitaria de la manifestaria de la versitaria de la mandar manafarte final de la mandar manafarte final de la mandar manafarte final de la mandar manifestoria de la della della manifestoria de la della della manifestoria de la della della manifestoria della della della manifestoria della	emental Deta	
A second	manifes or el produce posi interpropria primo del posi interpropria primo del posi interpropria primo del posi interpropria primo del managetti fuel conservante interpropria primo del posi interpropria primo del posi interproprimo del		Liters
A second se	manifes or el produce posi interpropria primo del posi interpropria primo del posi interpropria primo del posi interpropria primo del managetti fuel conservante interpropria primo del posi interpropria primo del posi interproprimo del	ementical Optica	Lines
A long of the A	In the second se	emential Data	
b Starting Letter Ag Lette allowering de series de se	manifes or all productions of al	emethal Deta	3
A subset of the A with the adverse of the adverse of the A with the adverse of the A with the	manifes or a large cade of the second	Internal Data	
An and a set of	manifes or all productions of al	Internal Data	3
A subset of the A subset of th	In the operation of the second	Internal Data	
A straight of a	In the operation of the second	Internal Data	

Submit genetic maps, association or population study data

Most submissions from journal requirement: Tree Genetics and Genomes

treegenesdb.org

#### User submissions

## Genetic maps, association or population studies

TGAD Accession Request Form		
	tic, phenotypic, and/or environmental data into the	
Genes database. Results of the association tests ctively available using a TGAD accession. Result	will be accepted as raw P-values and all files will be	
cavery available using a TGAD accession. Result cation of the associated manuscript.	can be released immediately or held until	
	TGAD Accession Request Form	
ecles ase fill in the fields below for data pertaining to Ple		
se till in the tields below for data pertaining to Per sur study includes multiple species, each set of sp	acies data will be collecte	
	Please fill in the fields below for data pertaining to Pleas teeds samples only. If your study includes multiple species, each set of species data will be collected one a	t a time.
netic Data		
s your study include Genetic data?	GPS Data	
	Select the statement that describes how GPS coordinate data was used in your study:	
	GPS coordinates were not collected in this study.	
enotype Data	Phenotype Data	
is your study include Phenotype?	Select the statement that describes how Phenotype data was used in your study:	
8	Defined Phenotype measurements were performed on each genotyped _	
vironmental Data	Environmental Data	
is your study include Environmental data?	Select the statement that describes how Environmental data was used in your study:	
	Environmental measurements were performed on each genotyped indiv -	
•	Association Results Data	
	Select the statement that best describes your Association results data:	
Co	tinue Single Marker -	
	Continue	
Note: The second	enveright that has arready	
The second secon	nenet to fin a grange party strange and grange party handlesses. Thistome strange without statement methods statement	
The spectra starts of the address of grants of the spectra start of the spectra starts o	nanati fai na anna Nangi wa ghang na Nangi wa ghang na Nangi wa ghang na Nangi wa ghang na na Nangi wa ghang na na	
A second	nenet to fin a grange party strange and grange party handlesses. Thistome strange without statement methods statement	
A second	nenet to fin a grange party strange and grange party handlesses. Thistome strange without statement methods statement	
<ul> <li>The second second</li></ul>	neneta da resuera na regi en de para de resuera da la construcción resuera	
A second	And the first of a series of the series of t	
In the part of the set of the advectory of the set o	neneta da resuera na regi en de para de resuera da la construcción resuera	
The Section of the Section of Sec	Add Supplemental Data For instruction on the acceptable formats for these fields, <u>cick.heep</u>	
Beneficial Control Contro	In the same many set of the sa	
A set of the set of a set	Adds Toppstate fields, click here Provide the comparison of the second	
	Add Supplemental Data Add Supplemental Data Add The Supplemental Data For Instructions on the acceptable formats for these fields, <u>citik base</u> Capacitan bala Part [fain.Adds 2] Part [	
The second se	Adds Toppstate fields, click here Provide the comparison of the second	
	Add Supplemental Data  Add Supplemental Data  Add Supplemental Data  Add Supplemental Data  Add Supplemental Control for these fields, cick base  Control for the supplemental Control for these fields, cick base  Control for the supplemental Control for these fields, cick base  Control for the supplemental Control for these fields, cick base  Control for the supplemental Control for these fields, cick base  Control for the supplemental Control for these fields, cick base  Control for the supplemental Control for these fields, cick base  Control for the supplemental Control for these fields, cick base  Control for the supplemental Control for these fields, cick base  Control for the supplemental	
A sector of the sector of	Add Supplemental Data Add Supplemental Data Add Supplemental Data Add Supplemental Data For instructions on the acceptable formats for these fields, click.here Cognectation late Add Supplemental Data For instructions on the acceptable formats for these fields, click.here Cognectation late Add Supplemental Cognet Generation	
	Add Supplemental Data  Add Supplemental Data  Add Supplemental Data  Add Supplemental Data  Add Supplemental Control for these fields, cick base  Control for the supplemental Control for these fields, cick base  Control for the supplemental Control for these fields, cick base  Control for the supplemental Control for these fields, cick base  Control for the supplemental Control for these fields, cick base  Control for the supplemental Control for these fields, cick base  Control for the supplemental Control for these fields, cick base  Control for the supplemental Control for these fields, cick base  Control for the supplemental Control for these fields, cick base  Control for the supplemental	
A sector of the sector of	Add Supplemental Data Add Supplemental Data Add Supplemental Data Add Supplemental Data For instructions on the acceptable formats for these fields, click.here Cognectation late Add Supplemental Data For instructions on the acceptable formats for these fields, click.here Cognectation late Add Supplemental Cognet Generation	
A sector of the sector of	And the for a server the server of the server and Add Supplemental Data For instructions on the acceptable formats for these fields, click here Cognization biol Add T [Ener. Advent ] Add D [Ener. Advent ] Add D [Ener. Advent ] For instructions on the acceptable formats for these fields, click here D [Point ] (Ener. Server ] (Ener. S	

0		ons is displayed below. nes, click here.			
Date	Accession	Paper Title	Species	Data Statistics	Data Files
8/5/2011	TGDR001	Association genetics of traits controlling lignin and cellulose biosynthesis in black cottonwood (Populus trichocarpa, Salicaceae) secondary xylem.	Populus trichocarpa	Total Sites: 1 Total Samples: 480 Total Gendypes: 419520 Total AFLP Markers: 0 Total SNP Markers: 0 Total SNP Markers: 0 Total SSR Markers: 0 Total SSR Markers: 0 Total Phenotypes: 1344 Total Environmentals (per sample): 0 Total Environmentals (per site): 0	Covariate Data (Population Struc Genotype Data (SNP) GPS Data Haplotype Data Phenotype Data Phenotype Definitions
9/25/2012	TGDR002	Astonishingly low genetic variation in Quercus acutissima, an important tree species in Satoyama, a traditional Japanese rural forest and agricultural landscape, revealed by chloroplast microsatellite markers	Quercus acutissima	Total Sites: 59 Total Samples: 2152 Total Genotypes: 12912 Total AFLP Markers: 0 Total SIN Markers: 0 Total cpSSR Markers: 6 Total SSR Markers: 0 Total Provident Site (per sample): 0 Total Environmentals (per sample): 0 Total Environmentals (per site): 0	Genotype Data (cpSSR) GPS Data Haplotype Data Supplemental Data
11/5/2012	TGDR003	Extensive selfing in an endangered population of Pinus parviflora var. parviflora var. (Pinaceae) in the Boso Hills, Japan	Pinus parviflora	Total Sites: 2 Total Samples: 116 Total Gendypes: 464 Total AFLP Markers: 0 Total RAPD Markers: 0 Total SRM Markers: 0 Total SRM Markers: 4 Total SRM Markers: 4 Total Phenotypes: 0 Total Environmentals (per sample): 0 Total Environmentals (per site): 0	Genotype Data (SSR) GPS Data Supplemental Data Supplemental Data Supplemental Data Supplemental Data
11/14/2012	TGDR004	Pollen dispersal and fine-scale spatial genetic structure of Dryobalanops lanceolata in a Bornean rain forest	Dryobalanops lanceolata	Total Sites: 1 Total Samples: 858 Total Genotypes: 13728 Total AFLP Markers: 0 Total SNP Markers: 0 Total SNP Markers: 0 Total CSNP Markers: 0	Environmental Metric Data Environmental Metric Definitions Genotype Data (SSR)

#### treegenesdb.org

#### User submissions

## Genetic maps, association or population studies

TGAD Accession Request Form		
	etic, phenotypic, and/or environmental data into the s will be accepted as raw P-values and all files will be	
ctively available using a TGAD accession. Result		
cation of the associated manuscript.		
ecles	TGAD Accession Request Form	
ase fill in the fields below for data pertaining to PI		
sur study includes multiple species, each set of s	pecies data will be collecte Please fill in the fields below for data pertaining to Please teeds samples only.	
	If your study includes multiple species, each set of species data will be collected one	at a time.
netic Data is your study include Genetic data?	GPS Data	
	Select the statement that describes how GPS coordinate data was used in your stud	r
8	GPS coordinates were not collected in this study.	
enotype Data	Phenotype Data	
is your study include Phenotype?	Select the statement that describes how Phenotype data was used in your study:	
8	Defined Phenotype measurements were performed on each genotyped -	
	Environmental Data	
vironmental Data s your study include Environmental data?	Select the statement that describes how Environmental data was used in your study:	_
	Environmental measurements were performed on each genotyped indiv -	
8	Association Results Data	
	Select the statement that best describes your Association results data	-
G	Single Marker -	
	Continue	
College Francesco Accessos e executor Ten Subergutorio Surtino administrario de protoci- tor statemento de la cinedación de composicionador de la consecutorio de la consecutorio de la serio a consecutorio de la consecutorio de la serio administrario de la consecutorio de la conse- tione a una consecutorio de la consecutorio de entre publicava por funcio de la consecutorio de la consecutorio de la consecutorio de la consecu- torio de la consecutorio de la consecutorio de la consecutorio de la consecutorio dela consecutorio de la consecutorio del la consecutorio del consecutorio de la consecutoria de la consecutorio de la consecutorio de la consecutorio de la consecutorio de la consecutorio de la consecutorio de la consecutorio de la consecutorio de la consecutorio de la consecutorio de la consecutorio	man fails and with provide the winnershift with the main cost for main structures cost of the main structures manualized fails fails elements	
The spectra bar is up to adhere up grand by a spectra bar is a spectra of the spectra bar is a spectra of the spectra bar is a spectra of the spectra bar is a spectra bar is spectra bar is a spectra bar is a sp		
The second secon		
The spectra of the sp	man Tana para para para bala para para para para para para para para	
The spectra of the second seco		
A sector of the sector of	The operation of the second of the operation of the second of the secon	
The Section of the Se	The second secon	
Beneficial of the set of the additional of the other set of the other	The second secon	
Beneficial of the set of a difference of the set o	Add Supplemental Data     Addr Guptof Tables     Addr Supplemental Data	
	The second and the se	
In the second	Add Supplemental Data     Constant on the acceptable formats for these fields, click here     Add Supplemental Data     For instructions on the acceptable formats for these fields, click here     Add Supplemental Data	
	The second and second	
In the second	The set of	
	The second and second	
A set of a set o	The second secon	
A set of a set o	The second secon	

#### Obtain TGDR accession number!

#### TreeGenes Data Repository

A listing of data submissions is displayed below. To submit data to TreeGenes, click here.

Date	Accession	Paper Title	Species	Data Statistics	Data Files
8/5/2011	TGDR001	Association genetics of traits controlling lignin and cellulose biosynthesis in black cottonwood (Populus trichocarpa, Salicaceae) secondary xylem.	Populus trichocarpa	Total Sites: 1 Total Samples: 480 Total Genotypes: 419520 Total AFLP Markers: 0 Total RAPD Markers: 0 Total SNP Markers: 874 Total cpSSR Markers: 0 Total SSR Markers: 0 Total SPhenotypes: 1344 Total Environmentals (per sample): 0 Total Environmentals (per sample): 0	Covariate Data (Population Structure) Genotype Data (SNP) GPS Data Haplotype Data Phenotype Data Phenotype Definitions
9/25/2012	TGDR002	Astonishingly low genetic variation in Quercus acutissima, an important tree species in Satoyama, a traditional Japanese rural forest and agricultural landscape, revealed by chloroplast microsatellite markers	Quercus acutissima	Total Sites: 59 Total Samples: 2152 Total Genotypes: 12912 Total AFLP Markers: 0 Total RAPD Markers: 0 Total SNP Markers: 0 Total SNR Markers: 6 Total SNR Markers: 0 Total Phenotypes: 0 Total Phenotypes: 0 Total Environmentals (per sample): 0 Total Environmentals (per site): 0	Genotype Data (cpSSR) GPS Data Haplotype Data Supplemental Data
11/5/2012	TGDR003	Extensive selfing in an endangered population of Pinus parviflora var. parviflora var. (Pinaceae) in the Boso Hills, Japan	Pinus parviflora	Total Sites: 2 Total Samples: 116 Total Genotypes: 464 Total AFLP Markers: 0 Total AFLP Markers: 0 Total SPM Markers: 0 Total SSR Markers: 0 Total SSR Markers: 4 Total Phenotypes: 0 Total Environmentals (per sample): 0 Total Environmentals (per site): 0	Genotype Data (SSR) GPS Data Supplemental Data Supplemental Data Supplemental Data Supplemental Data
11/14/2012	TGDR004	Pollen dispersal and fine-scale spatial genetic structure of Dryobalanops lanceolata in a Bornean rain forest	Dryobalanops lanceolata	Total Sites: 1 Total Samples: 858 Total Genotypes: 13728 Total AFLP Markers: 0 Total RAPD Markers: 0 Total STAP Markers: 0 Total GSSR Markers: 0	Environmental Metric Data Environmental Metric Definitions Genotype Data (SSR)

#### treegenesdb.org

#### User submissions

## Genetic maps, association or population studies

TGAD Accession Request Form			
This submission process will allow you to upload gener	c, phenotypic, and/or enviro		
TreeGenes database. Results of the association tests	vill be accepted as raw P-val		
collectively available using a TGAD accession. Results publication of the associated manuscript.	can be released immediatel		
			- V
Species			w w
Please fill in the fields below for data pertaining to Plin If your study includes multiple species, each set of spi			
Genetic Data			
Does your study include Genetic data?			
Yes			
Phenotype Data			
Does your study include Phenotype?			
Yes	*		
Environmental Data			
Does your study include Environmental data?			
Yes			
100			
Cor	tinue		
Ottalin Treadlering Accession Mideller			
	address and all recently		
the submitter with a Tree-Denies appendix whether I is the Tree-Denies database. This interface is internet	had will refer brick the max.		
with a monutacripit that will be submitted to a perime			
Fixes would like to submit the mapping films for a me term putterned, preser follow the term			
Prease entor preletinary internation social year eur	Lacquarie speak your		
Freese enter previously intendion stoud your war genete wasp encore the per investigation of the test remainstance and the terminiting of the tres can be			
Episoneg successful subressor, you will remove a surface on the third coder and value-has there go pate in our collection discuss to organities subre	construction and the set of		
Final interview the TreeConex database and the received term in the manuality retended for submit			
For more information about how to campion deal for teresticative	Prise Column		
S Freedom State Andrew			
AFTS STOLLERARS			
ADV1     Loui			
CT of Call, S. Justice , Secondary, B. ), Support	Add Suppler		-
Adhar Pesis Ioana Sentega C. Generation Imain an Journel D. Navas	For instructions on	the acceptable	formats for these
Publication Information	Organization Info		
<ul> <li>The Desire A partner distance distance of the owner of the</li> </ul>			
<ul> <li>South Stream</li> <li>Discussion</li> <li>South Stream</li> </ul>	Author Esken	Acident	
Distance (Mar	Paper Patient	e of population stru	man and accounting
3 decret ter			
Conductory	Supplement Genetic Type	map	
Copercision Committee Committee	Genetic Map		
			Concerning and Concerning
Subrat Internation	- lobe	esic_map_ecki_B	
	Organization Using	raity of California at	Davis
		ubmit Informatio	in .

#### Will be converted to Tripal module and made available to the community

		markers			
11/5/2012	TGDR003	Extensive selfing in an endangered population of Pinus parviflora var. parviflora (Pinaceae) in the Boso Hills, Japan	Pinus parviflora	Total Sites: 2 Total Samples: 116 Total Gendypes: 464 Total AFLP Markers: 0 Total SNP Markers: 0 Total SNP Markers: 0 Total SSR Markers: 0 Total SSR Markers: 4 Total Phenotypes: 0 Total Environmentals (per sample): 0 Total Environmentals (per sample): 0	Genotype Data (SSR) GPS Data Supplemental Data Supplemental Data Supplemental Data Supplemental Data
11/14/2012	TGDR004	Pollen dispersal and fine-scale spatial genetic structure of Dryobalanops lanceolata in a Bornean rain forest	Dryobalanops lanceolata	Total Sites: 1 Total Samples: 858 Total Genotypes: 13728 Total AFLP Markers: 0 Total RAPD Markers: 0 Total SNP Markers: 0	Environmental Metric Data Environmental Metric Definitions Genotype Data (SSR)

Data Files

er sample): 0

er sample): 0 er site): 0

er site): 0

Phenotype Data Phenotype Definitions

Genotype Data (cpSSR) GPS Data Haplotype Data Supplemental Data

Covariate Data (Population Structure) Genotype Data (SNP) GPS Data Haplotype Data

# TreeGenes Database: Data Access

treegenesdb.org

## Tools

- Existing viewers
- Custom development

## TreeGenes Database: Interfaces



# TreeGenes Database: Interfaces

#### treegenesdb.org

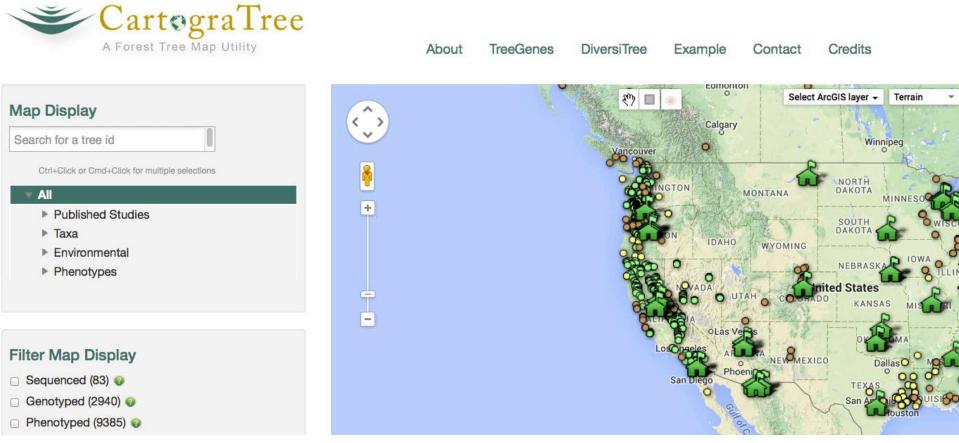
Cryptomeria japonica

#### **DiversiTree** – Bulk retrieval of resequencing data, genotypes, and phenotypes

Search	ies	<b>@</b>	•			cartogratree treegenes dendrome plone
		_				
+ Docu	mentation	?		tig_     copper ion binding   Al		
	(C. D. 1)		ID 📥	Contig_Name   0 11582	Species A Pinus taeda (Loblolly Pine)	🔄 🔄 🛞 sswap.info/lpc/la4d43d44-60b1-4c65-8ba0-a84d69b55a07 🛛 🗸 C 🔞 🕏 Coogle 🔍 🏠 🛱 🦑 💌 🚍
+EST	(GenBank)	?	1697 8140	0_11582	Pinus taeda (Lobioliy Pine) Pinus taeda (Lobioliy Pine)	📷 Most Visited 👻 门 Getting Started 🔝 Taraneh Shared 👻 🚸 CC student jobs!!
			8866	0_1829	Pinus taeda (Lobiolly Pine)	
+ Cont	ıg	?	9909	0_2414	Pinus taeda (Lobiolly Pine)	course into
FASTA S	equence:	<u> </u>	10957	0_3937	Pinus taeda (Lobiolly Pine)	sswap.info
	(Ex: GCATAG)		11256	0_4342	Pinus taeda (Loblolly Pine)	Simple Semantic Web Architecture and Protocol
Cont	g Name:		13739	0_7878	Pinus taeda (Loblolly Pine)	New pipeline
	(Ex: 0_8156)		14065	0_8291	Pinus taeda (Loblolly Pine)	
	_		14200	0_8419	Pinus taeda (Loblolly Pine)	
	GenBank cession:	<u></u>	14320	0_8531	Pinus taeda (Loblolly Pine)	TrecGenes
	(Ex: CBI25984)		15122	0_9288	Pinus taeda (Loblolly Pine)	
BLAST	GenBank	2	15266	0_9423	Pinus taeda (Loblolly Pine)	
	GI: (Ex: 270238748)		22822	2_1343	Pinus taeda (Loblolly Pine)	C Loui Date Set
RI	AST Hit:		30763	2_8946	Pinus taeda (Loblolly Pine)	Input Data Set
			cg47012	CL1367Contig1	Pinus taeda (Loblolly Pine)	TreeGenes Contraction Data Set URI: http://sswap.info/ipc/rrg?token=4ce1fc5a-128e 48fa-9b93-4dc7ea771869
	(Ex: p-type ATPase	:)	cg48716	CL2662Contig1	Pinus taeda (Loblolly Pine)	Config Data Set UKI: http://sswap.info/pp/rgr.token=4ce1rc3a-128e48ita 9095-44c/ca7/1869 Produced by: http://sswap.dendrome.ucdavis.edu/resources/contigService/ContigService
C	O Term: copper ion bind	ling	cg49217	CL3147Contig1	Pinus taeda (Loblolly Pine)	
	(Ex: Heat shock)		cg50235	CL4447Contig1	Pinus taeda (Loblolly Pine)	copper ion binding Yes Yes
InterP	ro Term:		cg50336	CL4575Contig1	Pinus taeda (Loblolly Pine)	copper ion binding Yes Yes
	(Ex: Lipid)		cg50472	CL4726Contig1	Pinus taeda (Loblolly Pine)	copper ion binding Yes Yes
			cg50688	CL569Contig1	Pinus taeda (Loblolly Pine)	copper ion binding Yes Yes
	Enzyme:					Sswap.info
	(Ex: ATP)					
Mark al	species of interest:					
	Select: ALL   NONE					
	Abies alba					
	Abies concolor					
☑	Calocedrus decurrens					
	Cedrus deodara					

## TreeGenes Database: Interfaces

#### treegenesdb.org



Providing context to geo-referenced data
 Data from TreeGenes, WorldClim, Ameriflux, TRY-db

## TreeGenes Database: Transition to Tripal

treegenesdb.org

### **Transition to Tripal and Chado**

## Challenges

- Ontologies
- Standardizing for the first time
  - 25 years of custom schemas!

# Transition to Tripal

treegenesdb.org

## **Transition to Tripal and Chado**

### Advantages

- Ontologies
- Save time!
- Connect data with other databases (Tripal Gateway)
- Improve analytical capabilities (Tripal Gateway)
  - Our contribution: analytical workflows
- Review & clean up database

## TreeGenes Database: Transition to Tripal

treegenesdb.org

#### Outcomes

–Expanded datasets: access Hardwood Genomics, GDR, TreeGenes from one location

-Expanded analytical, HPC resources

•Improve CartograTree capabilities

-New Tripal module: user submission

# TreeGenes Database: Team

#### treegenesdb.org

#### tg-help@ucdavis.edu

Project Leads Jill Wegrzyn *University of Connecticut* David Neale

Lead Database Administrator Emily Grau P0322 Development TeamSteven A Demurjian JrP0383Hans Vasquez-Gross

Advising Damian Gessler Semantic Options

TreeGenes
 TreeGenes
 Database
 Database
 UCCONNECTICUT

