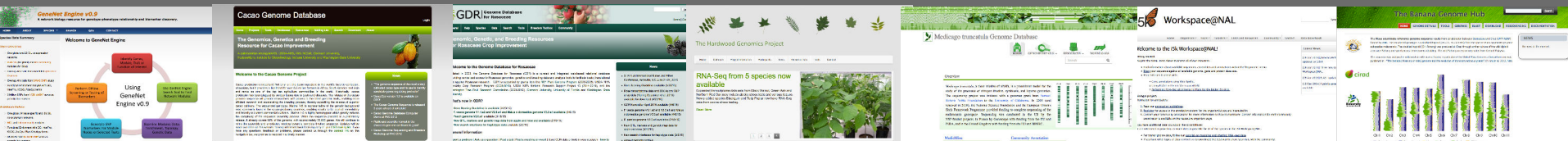


# Integrating Galaxy and Tripal

## Cyberinfrastructure for the Genome Community Database



tripal.info

Emily Grau  
Department of Ecology and Evolutionary Biology  
University of Connecticut

# Tripal Gateway

tripal.info



- NSF DIBBS, funded 2014
- Lead: Stephen Ficklin at WSU



## Goals



- Improve data transfer speed
- Make it easier to share data across community databases
- Provide NGS analysis tools in Tripal in collaboration with Galaxy



DIBBS

# Overview

tripal.info

## Background

- Databases involved
- What is Tripal?

## New development

- What is Tripal Gateway?
- Current Development
- Expected Outcomes



# Tripal Gateway Pilot Implementation

tripal.info

***For pilot, focus on 2 research communities:***

## **Legume community databases**

- Cool Season Food Legume
- KnowPulse
- Legume Information System
- PeanutBase
- *Medicago truncatula* Genome

**Tree community databases...**

# Tree Community Databases

tripal.info



## Genome Database for Rosaceae

- Economically important edible fruits
- >500 species

## Hardwood Genomics Project

- Hardwood forest trees (ecological and economic importance)
- ~16 species



## TreeGenes

- Forest trees (historic focus on conifers)
- >1700 species

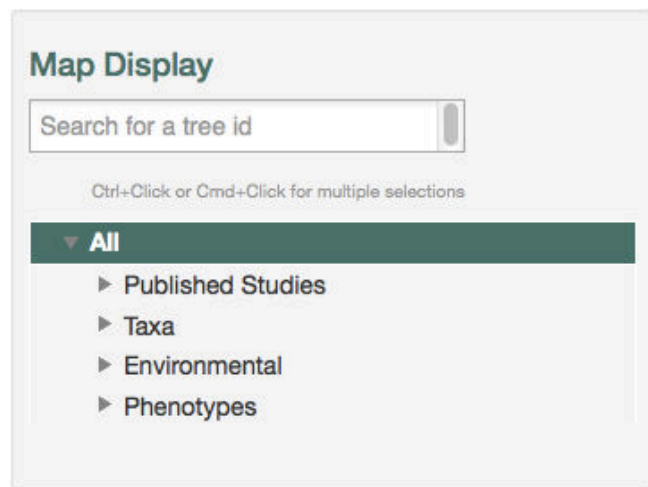


**User base** for the databases includes researchers, breeders, industry



# TreeGenes: CartograTree

tripal.info



- Providing context to geo-referenced data
  - Genotype, phenotype, and environment
- Enable analysis GWAS (e.g. TASSEL) via web services

## Limitations

- Response time
- Data limitations
- Limited analytical tools

# Why Trees?

tripal.info

## Special Challenges

- Unique characteristics, biological questions
  - Integration with environment
  - Additional important datatypes (phenotype, environmental)
- **Broad** databases
  - Huge diversity of species (100s of species held)
  - Many without reference genomes
  - Result: fewer standards on other genetic objects; poor integration

## Common needs

- GWAS, synteny, RNA-Seq, variant discovery

# Why Trees?

tripal.info

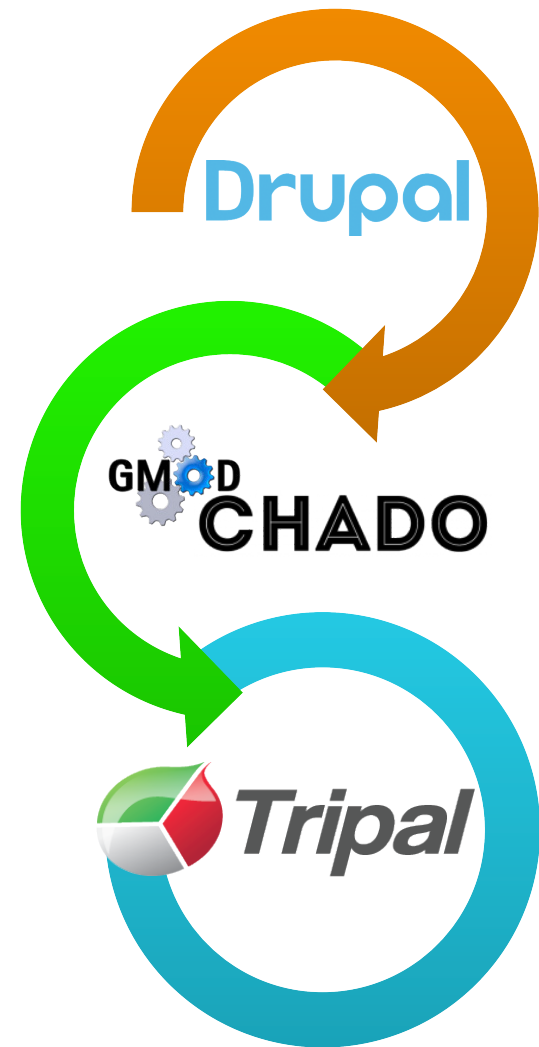
## Special Challenges

- Unique characteristics, biological questions
- Additional datatypes
- Huge diversity of species
- Few reference genomes

- **Tripal helps to store data in ontological manner**
- **Tools in development can benefit tree community, applicable to many other research communities**

# What is Tripal?

tripal.info



## Drupal

- Open source content management system (CMS)
- Active developer community
- 1000s of modules available

## Chado

- Relational database schema for GMOD project

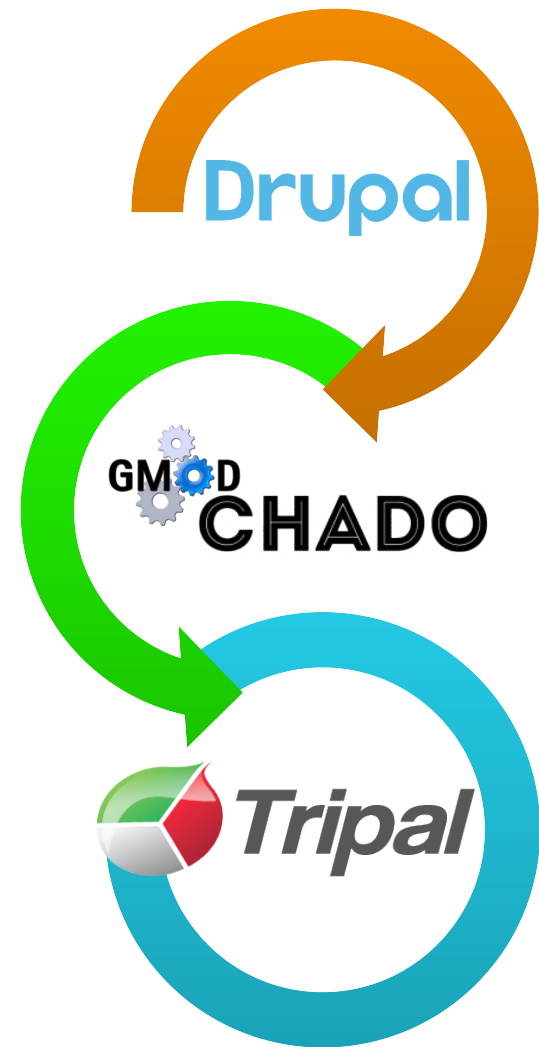
**Tripal is a suite of Drupal modules**

# Tripal

tripal.info

## Developed to assist community databases with few resources

- Out of the box functionality
  - Reduce \$, resources needed
- Flexible and customizable
- Open source
  - Custom development (shareable)
- In use by > 60 projects (July 2015)



# Challenges for Community Databases

tripal.info

## **Tripal was developed to help community databases address challenges:**

- How to maintain large datasets with limited technical support?
- How to implement existing community standards?

## **Tripal Gateway will address remaining challenges:**

- How to facilitate data sharing – quickly!
- How to provide analytical tools for next-generation data sets?

# Tripal Gateway: 3 Modules

tripal.info



DIBBS

- **Tripal Galaxy**
  - Integrate Galaxy workflows within a Tripal site
- **Tripal Exchange**
  - Allow cross-database querying
- **Tripal SDN**
  - Software defined networking to optimize data flow mechanisms
- *Modules will work together to improve user experience.*

# Tripal Galaxy

tripal.info



## Why integrate a GUI?

- Streamline process for users
  - Less confusing
  - Integrated user authorization
- Add value to community database
  - Provide analytical tools tailored to community



# Tripal Galaxy

tripal.info



## User Experience

- Users will be able to select a Galaxy workflow from list
- Users can use data from:
  - Tripal sites via Tripal Exchange module
  - Public repositories (e.g. GenBank)
  - Private data
- Module will handle job submission

# Tripal Galaxy

tripal.info



## Development Plans

- Work begins late 2015
- PHP bindings for Galaxy API to be developed
  - Usable outside of Tripal
- Tripal Galaxy stages data on remote computational resource (usegalaxy.org, local Galaxy instance, etc)



# Tripal Galaxy

tripal.info



## Development Plans

- Analytical workflows constructed in Galaxy, made available in Tripal
  - Workflows will be adapted/created by tree databases & others
  - Workflows & tutorials will be shareable in/outside of Tripal



# Tripal Galaxy

tripal.info



Workflows (to be adapted or created):

– **DNA (genome)**

- Re-sequencing alignment
- Variant discovery (against reference and between samples)
- Syntenic block analysis for comparative genomics

– **RNA (transcriptome)**

- Transcriptome assembly
- Variant discovery (against reference and between samples)
- Gene co-expression network construction
- MiRNA analysis

– **Trait association**

- Genome-wide association
- Prediction of functional genetic variants



DIBBS

# Tripal Exchange

tripal.info



- Provide a RESTful web service framework
- Tripal Galaxy workflows will be able to access all data across Tripal sites for analysis



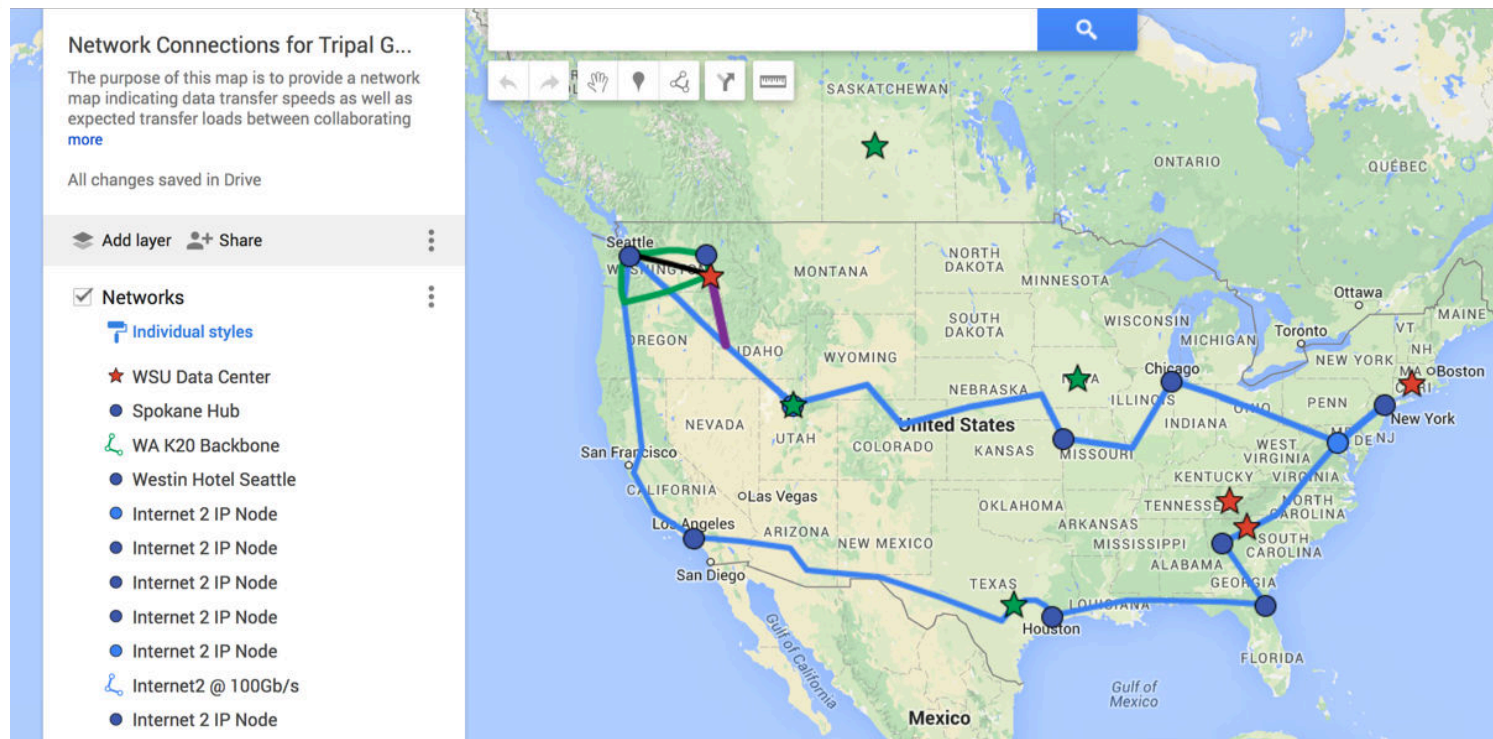
# Tripal SDN

tripal.info



Alex Feltus, &  
KC Wang

- Transfer/visualize data quickly through software-defined networking (SDN) pipes
- Hardware recommendations for optimal transfer speeds will be made



# TreeGenes : CartograTree

tripal.info



[About](#)

[TreeGenes](#)

[DiversiTree](#)

[Example](#)

[Contact](#)

[Credits](#)

**Tripal Gateway** will allow us to:

- Reduce query time (Tripal SDN)
- Draw data from GDR, Hardwood Genomics (Tripal Exchange)
- Offer more analytical pipelines to users (Tripal Galaxy)

# Tripal Gateway: Benefits

tripal.info



## Non-Tripal Users

- Workflows & tutorials usable outside of Tripal



## Non-Tripal Admins

- PHP bindings for Galaxy
- Tripal Exchange Web Services usable outside of Tripal



DIBBS

# Tripal Gateway: Benefits

tripal.info



## Tripal Users

- Faster data transfer (Tripal SDN)
- Seamlessly fetch data across databases (Tripal Exchange)
- Seamless use Galaxy workflows (Tripal Galaxy)



## Tripal Admins

- All of the above!
- Hardware recommendations to optimize data transfer speeds



DIBBS

# Tripal Gateway

tripal.info



DIBBS

## We want your input!

Monthly meetings (1<sup>st</sup> Tuesday)

Tripal Mailing lists:

[lists.sourceforge.net/lists/listinfo/gmod-tripal](https://lists.sourceforge.net/lists/listinfo/gmod-tripal)

[lists.sourceforge.net/lists/listinfo/gmod-tripal-devel](https://lists.sourceforge.net/lists/listinfo/gmod-tripal-devel)

# Acknowledgments

**tripal.info**

## **Tripal**

Stephen Ficklin  
Lacey Sanderson  
Chun-Huai Cheng  
Connor Wytko  
Brian Soto

## **Software Defined Networking**

Alex Feltus  
Kuangching Wang  
Nick Watts

## **Hardwod Genomics Project**

Margaret Staton  
Nathan Henry

## **Genome Database for Rosaceae**

Doreen Main  
Sook Jung  
Jodi Humann  
Taein Lee

## **TreeGenes**

Jill Wegrzyn  
Emily Grau  
Steve Demurjian

## **Semantic Web**

Damian Gessler



**DIBBS**



THE UNIVERSITY  
OF ARIZONA

