

# Tripal elasticsearch module

Ming Chen @ Staton Lab





A suite of



modules for construction of biological web databases

<http://tripal.info>

Home Download Support Extensions API About Demo Site Log In

**Quick Links**

- v2.x User's Guide
  - Quick Docker install
  - Download and install
  - Upgrade instructions
  - Developer's handbook
- Other documentation
- Tripal @ GitHub.com
- Tripal @ drupal.org

**Tweets by @TripalProject**

The Tripal Project @TripalProject Gene Ontology Consortium meeting 2017 @ Oregon State June 1-5 GOC meeting, Noctua gene annotation, Reactome Pathway sites.google.com/view/goc2017 25 Jan

The Tripal Project Retweeted @JillWegrzyn Accepting Postdoc applications for @TreeGenes #BigData #Bioinformatics Association Genetics

**Tripal**

Tripal is a toolkit for construction of online biological (genetics, genomics, breeding, etc), community database, and is a member of the GMOD family of tools. Tripal provides by default integration with the GMOD Chado database schema and Drupal, a popular Content Management Systems (CMS). Tripal is licensed under the GNU General Public License version 2.

**Objectives**

- Provide a framework for those with genomic, genetic and breeding data for creation of an online site for display, search and visualization.
- To use community-derived standards and ontologies to facility continuity between sites which in

**Sites Using Tripal**

Genome Database for Vaccinium

[View more]

**News**

**TRIPAL: Meeting & Hackathon**  
01/03/2017 - 23:42  
Where: Plant & Animal Genome, San Diego, CA  
Town & Country in Royal Palm Salon 4

<https://github.com/tripal>

**Tripal**  
An open-source toolkit for construction of genomic, genetic, breeding and biological databases.  
<http://tripal.info>

**Repositories** People 24 Teams 4 Projects 0

Search repositories... Type: All Language: All Customize pinned repositories **New**

**tripal\_elasticsearch**  
● PHP 3 Updated 8 hours ago

**tripal\_galaxy**  
This module is for integration of Tripal and a remote Galaxy Instance  
● PHP Updated 15 hours ago

**tripal**  
The Tripal "core" package, a suite of Drupal modules for creating biological (genomic, genetic, breeding) websites. Visit the Tripal homepage at <http://tripal.info> for documentation, support, and other information. Please use the Drupal project's issue queue at <https://www.drupal.org/project/issues/tripal> for bug reports and requests.  
● PLpgsql ★ 9 9 Updated 18 hours ago

**Top languages**  
● PHP ● JavaScript ● PLpgsql

**People** 24 >

Main Lab Bioinformatics



**City of Chicago** The City of Chicago's Official Site

English | Español | 中文 | Polski | العربية

Keyword  Search

Home City Services People We Serve Programs & Initiatives Chicago Government About Chicago

**City Services**

- I Want To...
- Apply For
- Check Status Of
- Find/Get
- Pay For/Buy
- Register
- Report/File
- Request
- Sign up for/Volunteer

**APPLY BY JANUARY 31, 2017**  
MAKE 72K WITHIN 18 MONTHS  
**BE THE CHANGE** [www.bethechangecity.com](http://www.bethechangecity.com)

**Building on Burnham:** A Plan to Bring the "City in a Garden" to Every Chicago Neighborhood

Greetings from Mayor Rahm Emanuel

Multimedia

Connect with the City of Chicago via Social Media

The Chicago Park District YouTube Channel

City of Chicago TV

University of OXFORD Home

ADMISSIONS | RESEARCH | NEWS & EVENTS | ABOUT

Colleges | Divisions & departments | Staff | Oxford students | Alumni | Visitors | Local community

**REVIEW 2015-16**

Cookies to help give you the best experience on our website. By continuing without changing your cookie settings, we agree to this. Please read our [cookie policy](#) to find out more.

**DISMISS**

**BMJ**

WHO WE ARE WHAT WE DO PRODUCTS & SERVICES NEWSROOM WORK AT BMJ

Access The BMJ

First edition of PACK Adult Global in eBook launches.

The Practical Approach to Care Kit provides a generic 'framework' that can be customised to help primary healthcare systems in individual countries or states.

Purchase your copy | [pack.bmj.com](http://pack.bmj.com)

**Drupal**

**THE BEATLES**

THE BEATLES LOVE EXPLORE ALBUMS SONGS NEWS STORE ▶ PLAY MUSIC

LIVE AT THE HOLLYWOOD BOWL

A RON HOWARD FILM

EIGHT DAYS A WEEK

THE TOURING YEARS

**THE BEATLES**



# Problems that tripal-based databases

- Lacking site wide indexing/search
- Can't index all data from the chado database



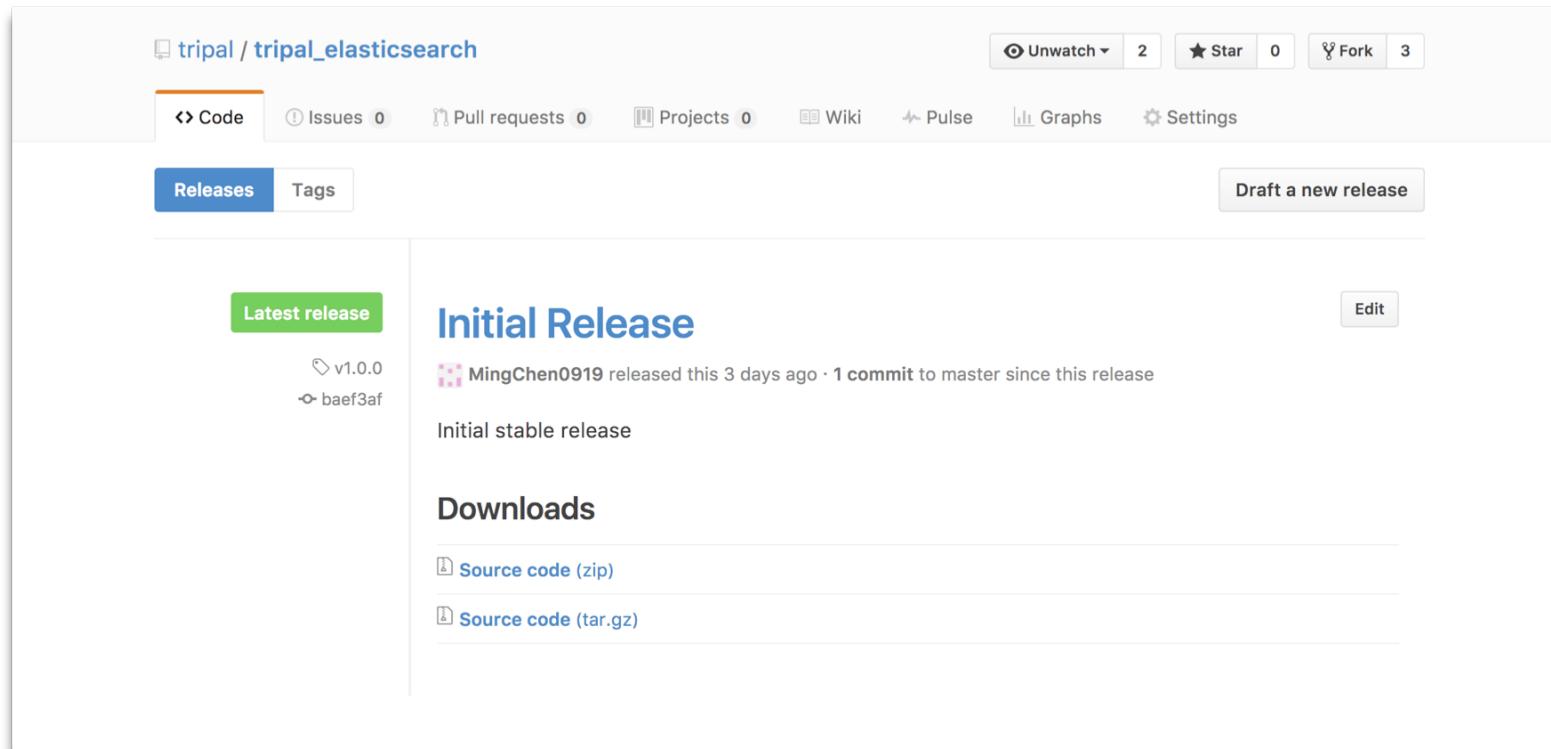


We need a search engine that provides efficient indexing and search, like  elastic

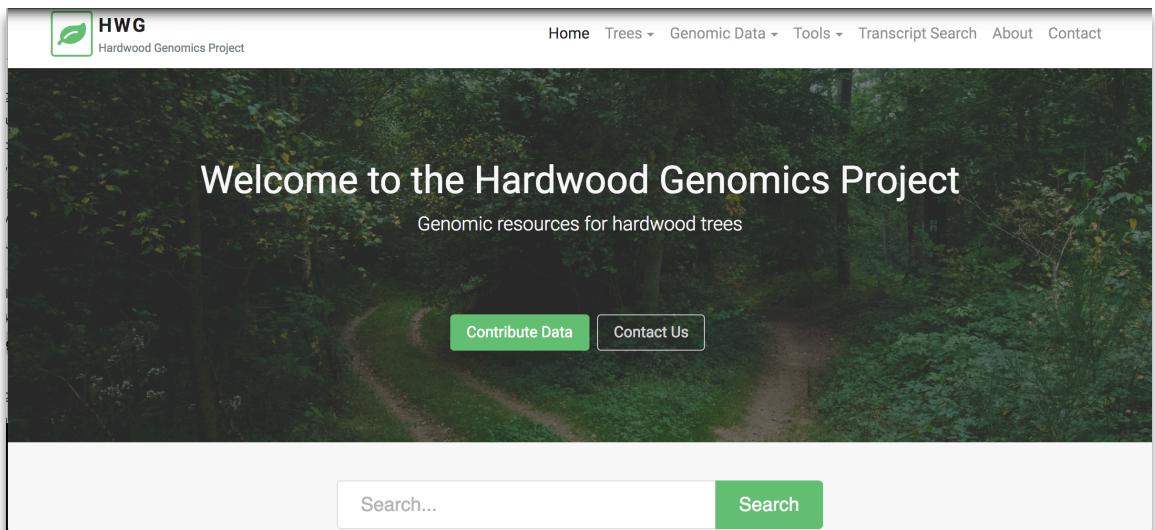
# Why elastic

- Indexed data becomes searchable immediately, even the indexing process is still ongoing.
- Hundreds of search APIs that provides fast, incisive search against large volumes of data
- Broadly distributable and highly scalable:
  - An elasticsearch cluster can consist of nodes that are hosted at different servers
  - You can add or remove nodes to change the size of elasticsearch cluster

# Tripal elasticsearch module on github



# Implementation on the Hardwood Genomics Database (HGD)



## Transcripts Search

Any

Sequence Unique Name

BLAST match description

**Search:** clicking on the search button without keywords input will return all available records.

- Can index every single page
- Any page content change will be automatically detected and re-indexed
- Can build interface form for any tables from the database
- Allows connection to both local and remote cluster

- Wildcard search (\*): genome\* sequence -> genome, genomic, genomics ...
- Fuzzy search (~): sequeeence~ -> sequence
- Regular expression search (/.../): /transcriptom[a-z]+/ -> transcriptome ....
- Boolean operators (+/-): +"green ash" +transcriptome -genome
- AND, OR, NOT, phrase search ("") and parenthesis to group searches
  - "heat stress" AND ("Castanea mollissima" OR "green ash") NOT "heat shock"



University of Tennessee

- Meg Staton
- Nathan Henry
- Abdullah Almsaeed
- Michelle Zhou



The Tripal Community

- Stephen Ficklin (Washington State University)
- Lacey Anne Sanderson (University of Saskatchewan)
- Anthony Bretaudreau (INRIA/Irisa)
- Sudhansu Dash (NCGR)