

Tripal: an open source toolkit for building genomic and genetic data websites and databases

Stephen Ficklin¹, Lacey Anne Sanderson², Margaret Staton^{3*}, Chun Huai Cheng¹, Sook Jung¹, Kirstin Bett², Doreen Main¹

¹ Department of Horticulture, Washington State University

² College of Agriculture and Bioresources, University of Saskatchewan

³ Department of Entomology and Plant Pathology, University of Tennessee

* mstaton1@utk.edu

URL: <http://tripal.info/>

URL (code): <https://drupal.org/project/1337878/git-instructions>

License: GNU General Public License, Version 2

Community genomic databases fulfill a critical need by offering curated and mission-specific information to targeted audiences with shared basic and applied research goals. Tripal was created in response to the need for an open source, extensible software system to support and standardize the website development projects of diverse scientific communities with growing sequence resources and uniquely interlinked datasets. Tripal provides data pages and search tools for taxonomies, genomic sequences, genetic markers and maps, stocks, cultivars, DNA or clone libraries, and publications stored in Chado¹, the standard relational database schema for biological information. Tripal is written in PHP and provides an Application Programmers Interface (API) that allows other developers to extend the core modules and create new modules. As a result, extension modules have been developed for parsing and uploading data from common computational analyses of sequence data such as BLAST², InterProScan³, KEGG⁴, and blast2GO⁵. Tripal was developed for use with the popular open-source content management system Drupal⁶, a PHP-based platform used to power millions of websites and applications worldwide. Drupal empowers non-technical users to easily add content and functionality without the need for programming, including writing news articles, announcements or blogs as well as designing the layout and content of new pages. This robust underlying structure is leveraged in Tripal to allow curators or community members to upload different types of data through an intuitive web interface and to create new ways of visualizing heterogeneous datasets. Drupal also provides built-in user management and content versioning, an ideal base for constructing robust community annotation capabilities. By bridging Chado and Drupal, Tripal marries the power of a biological data storage schema with a web development platform to decrease the cost and time associated with development of genomic, genetic and breeding databases for diverse biological research communities. Tripal is in use or being implemented by at least 24 different databases including the Genome Database for Rosaceae⁷, CottonGen⁸, the Hardwoods Genomic Database⁹, and KnowPulse¹⁰. Future development is focusing on cross-site communication, adoption of community driven data standards, and integrated “big data” analysis.

1. http://gmod.org/wiki/Chado_-_Getting_Started

2. <http://blast.ncbi.nlm.nih.gov/Blast.cgi>

3. <http://www.ebi.ac.uk/Tools/pfa/iprscan>

4. <http://www.genome.jp/kegg>

5. <http://www.blast2go.com>

6. <https://drupal.org>

7. <http://www.rosaceae.org>

8. <http://www.cottongen.org>

9. <http://www.hardwoodgenomics.org>

10. <http://knowpulse.usask.ca>