

Methods for collecting, integrating, and displaying complex genetic data, using the Legume Information System and PeanutBase

Andrew D. Farmer¹, Scott Kalberer², Jugpreet Singh³, Ethalinda K.S. Cannon⁴, Pooja Umale¹, Hrishikesh Lokhande¹, Alan Cleary¹, Nathan T. Weeks², Vijay Karingula⁵, Sudhansu Dash⁶, Steven B. Cannon²

¹National Center for Genome Resources, Santa Fe, New Mexico, USA

²USDA – Agricultural Research Service, Corn Insects and Crop Genomics Research Unit, Iowa State University, Ames, Iowa, USA

³ORISE Fellow, USDA-Agricultural Research Service, Corn Insects and Crop Genetics Research Unit, Ames, IA, USA

⁴Iowa State University, Electrical and Computer Engineering, Ames, IA, USA

⁵Iowa State University, Management Information Systems, Ames, IA, USA

⁶Virtual Reality Application Center, Iowa State University, Ames, IA, USA

We will describe the Legume Information System (LIS, at <http://legumeinfo.org>) and PeanutBase (at <http://peanutbase.org>). Together, these sites (developed by the same USDA-ARS and NCGR teams) have many new features, spanning eight sequenced genomes in the legume family. We will briefly describe features, including search methods, gene family viewers, and genomic context visualizations – and will particularly focus on methods that we are using to collect, integrate, and display complex genetic data (QTLs, traits, maps, associated publications, etc.). The goal is to integrate many genetic studies with respect to common reference maps. Although this data curatorial aspect of science can be a chore, it is a very important part of making research available and useable – and is a critical function of genomic databases. We are using several approaches to make the problem more tractable. We are using spreadsheet templates for data collection, ontologies (term dictionaries) for trait descriptions, common schemas and tools for loading and displaying the data, and sharing of templates to help distribute the curatorial load. For storage and display, we are using Chado and Tripal, with a custom QTL display module; and CMap for map display. Several other Chado/Tripal projects are using similar approaches, raising the possibility of wider adoption of standard methods and more seamless interconnections between various genomic database projects.