

Abstract for **5th International Biocuration Conference**: April 2-4, 2012, Washington DC

<http://pir.georgetown.edu/biocuration2012.html/instructions.html>

Category for abstracts: Talk or Poster

Selected the following relevant topic sessions:

1. Ontologies, standards and best practices, including gold standard datasets
2. Community annotation and Wikis
3. Genomics and metagenomics data curation
4. High-throughput data (focus on NGS and MS data) curation and presentation.

Keywords: ontology, genomics data, curation, annotation, plant

Laurel Cooper¹, Mary Schaeffer², Ramona Walls³ Justin L. Elser¹, Justin Preece¹, Barry Smith⁴, Chris Mungall⁵, Marie A. Gandolfo⁶, Dennis Wm. Stevenson³ and Pankaj Jaiswal¹

(1) Oregon State University, Corvallis, OR, (2) USDA-ARS and University of Missouri, Columbia, MO, (3) New York Botanical Garden, Bronx, NY, (4) Department of Philosophy, University at Buffalo, Buffalo, NY, (5) Lawrence Berkeley National Lab, Berkeley, CA, (6) Department of Plant Biology, Cornell University, Ithaca, NY

Title: Annotating the Maize B73 Gene Expression Atlas – A Plant Ontology Use Case for Genomics Data Curation

The Plant Ontology (www.plantontology.org) is a structured vocabulary and database resource for all plant scientists that links plant anatomy, morphology, and development to the rapidly expanding field of plant genomics. The primary purpose of the PO is to facilitate cross-database querying and to foster consistent use of vocabularies in annotation. An essential feature of the PO is the set of freely accessible web links from terms to associated annotations, which are structure- or development-specific genes, proteins, and phenotypes sourced from numerous plant genomics datasets. In collaboration with MaizeGDB (www.maizegdb.org), we have recently added approximately 1.5 million new associations between maize (*Zea mays*) gene models and Plant Ontology terms. These associations are based on a large NimbleGen microarray data set profiling genome-wide transcription patterns in 60 tissues, representing 11 distinct organs over the life cycle of a maize plant of the inbred line B73 (Sekhon, et al, *Plant Journal*, 2011). The microarray data was associated with ~35,000 maize gene models developed from the recent sequencing of its genome and updated to the current assembly, B73 RefGen_v2, as a collaboration between MaizeGDB and PLEXdb (www.plexdb.org). The curation path used to associate the PO terms to the maize gene atlas began at MaizeGDB, by mapping each microarray tissue sample to Plant

Ontology terms for anatomy, growth and development. Although many of the necessary maize-specific terms already existed in the PO (largely from the early participation of MaizeGDB in the PO project), some minor clarification of term definitions and hierarchies at the PO were needed. These were primarily for the developmental stage terms, which required direct annotation. To ensure quality annotations of these data, curators at the PO and MaizeGDB undertook a highly interactive review process, and MaizeGDB was updated to reflect concurrent changes at PO in the representation of distinctive aspects of Poaceae and *Zea* reproductive structures. PO association files were prepared in gaf 2.0 format (www.geneontology.org) and further enhanced by the inclusion of classical gene names, mapped by CoGe (www.genomevolution.org/CoGe/). The maize gene atlas associations were made public in the Plant Ontology Release #16 in October, 2011. They are available for download, and can be viewed in various browser modes, both at the PO and at MaizeGDB. The addition of the maize gene atlas annotations to the PO represents an example of how ontologies provide access to large genomics data sets. Currently, the PO includes over 2 million such annotations from 17 species associated with over 1,300 terms. Other recent additions include annotations to cotton (*Gossypium*) and the moss *Physcomitrella patens*, with plans for the future inclusion of grape (*Vitis*) and potato (*Solanum*). The PO is a valuable resource for both research and teaching that can be used as a guide to plant structures and growth and developmental landmarks in life cycles of plants across many taxa.