Using the Plant Ontology to improve the onteroperability of genomic and phenomic data sets

To be presented at **Plant Genomes & Biotechnology: from Genes to Network** at Cold Spring Harbor Laboratory, November 30 - December 3, 2011

Ramona L Walls[1], Laurel Cooper[2], Justin Elser[2], Justin Preece[2], Christopher J Mungall[3], Maria A Gandolofo[4], Barry Smith[5], Dennis W Stevenson[1], Pankaj Jaiswal[2]

[1]New York Botanical Garden, Science Department, Bronx, NY, 10458, [2]Oregon State University, Deptartment of Botany and Plant Pathology, Corvallis, OR, 97331-2902, [3]Lawerence Berkeley National Laboratory, Berkeley Bioinformatics Ontology Project, Berkeley, CA, 94720, [4] Cornell University, Department of Plant Biology, Ithaca, NY, 14853, [5]University at Buffalo, Department of Philosophy, Buffalo, NY, 14260

The Plant Ontology (PO: http://plantontology.org) is a structured vocabulary (ontology) consisting of terms, attributes, and relations that describe anatomy, morphology, and development stages of green plants. In addition, the PO provides access to genes and phenotypes that have been associated with ontology terms via the annotation of samples from specific tissues and development stages. The PO is an essential, powerful tool for the annotation of diverse gene-expression and phenotype data sets that can be used to assess the similarity between genes of inter- or intra-specific origin and to explore homologies among plant structures. The PO facilitates computational reasoning, based on ontological relationships and biological context, allowing researchers to probe the complex relationships among data sets for gene expression, phenotypes, gene-gene interactions, and molecular functions (via the Gene Ontology). For example, the logical definitions and relationships in the PO can be used to deduce that 'petal' in dicots and 'lemma' in grasses are both subtypes of 'phyllome' (leaf-like structures) and that both are associated with 'flower' ('petal' is part_of 'flower' and 'lemma' is part_of 'inflorescence' which has_part 'flower'). Researchers can use the PO annotation data to compare the expression patterns of orthologous genes in these structures in maize and *Arabidopsis*, or to determine if similar phenotypes in the two structures are linked to orthologous genes. Currently, the PO includes over 2 million annotations from maize, *Arabidopsis*, strawberry, rice, solanaceous crops (such as tomato), and the moss *Physcomitrella patens*. These annotations are associated with over 1,400 ontology terms. Almost 400 new anatomical terms have been added to the PO recently, to enhance the framework for cross-species comparisons and accommodate work in future agricultural models such as *Musa* and *Eucalyptus*. Approximately 80 new terms were added specifically for non-vascular plants, with an emphasis on those needed to describe gene expression in *P. patens*. In this presentation, we will provide an overview of the Plant Ontology and its resources and present a pilot study comparing inter-specific gene expression profiles, based on the orthology of genes and ontological relations among plant structures.