AN EXTENSION OF THE PLANT ONTOLOGY PROJECT SUPPORTING WOOD ANATOMY AND DEVELOPMENT RESEARCH

Frederic Lens[1]*, **Laurel Cooper**[2], **Maria Alejandra Gandolfo**[3],
Andrew Groover[4], **Pankaj Jaiswal**[2], **Barbara Lachenbruch**[5], **Rachel Spicer**[6],
Margaret E. Staton[7], **Dennis W. Stevenson**[8], **Ramona L. Walls**[8] and
Jill Wegrzyn[9]

A wealth of information on plant anatomy and morphology is available in the current and historical literature, and molecular biologists are producing massive amounts of transcriptome and genome data that can be used to gain better insights into the development, evolution, ecology, and physiological function of plant anatomical attributes. Integrating anatomical and molecular data sets is of major importance to the field of wood science, but this is often hampered by the lack of a standardized, controlled vocabulary that allows for cross-referencing among disparate data types. One approach to overcome this obstacle is through the annotation of data using a common controlled vocabulary or "ontology" (Ashburner *et al*. 2000; Smith *et al*. 2007). An ontology is a way of representing knowledge in a given domain that includes a set of terms to describe the classes in that domain, as well as the relationships among terms. Each term can be associated with an array of data such as names, definitions, identification numbers, and genes involved. Ontologies are fundamental for unifying diverse terminologies and are increasingly used by scientists, philosophers, the military and online web search engines. In an ontology, terms are carefully defined, allowing a wide array of researchers to (1) use terms consistently in scientific publications or standardized handbooks on quality/trait evaluations, and (2) search for and integrate data linked to these terms in anatomical, genetic, genomic, and other types of biological databases.

The *Plant Ontology* (PO, www.plantontology.org) is a structured vocabulary and database resource that links plant anatomy and development to gene expression and phenotypic datasets from all areas of plant biology (Jaiswal *et al*. 2005; Avraham *et al*. 2008). The *Wood Ontology* project (http://wiki.plantontology.org/index.php/ Wood_anatomy_ontology_meeting,_2012_at_NYBG,_agenda), which was recently initiated during the *Wood Ontology Workshop* at the New York Botanical Garden on

1) Netherlands Centre for Biodiversity Naturalis, P.O. Box 9514, 2300 RA Leiden, The Netherlands.
2) Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR 97331, USA.
3) L. H. Bailey Hortorium, Department of Plant Biology, Cornell University, Ithaca, NY 14853, USA.
4) Institute of Forest Genetics, US Forest Service, Davis, CA 95618, USA.
5) Department of Forest Ecosystems & Society, Oregon State University, Corvallis, OR 97331, USA.
6) Department of Botany, Connecticut College, New London, CT O6320, USA.
7) Clemson University Genomics Institute, Clemson University, 51 New Cherry Street, Clemson, SC 29634, USA.
8) New York Botanical Garden, Bronx, New York 10458, USA.
9) Department of Plant Sciences, Davis, CA 95616, USA.
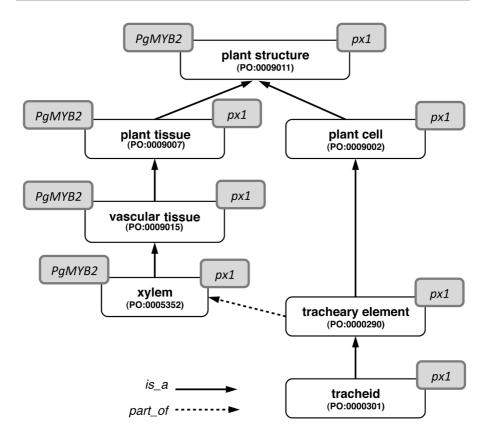*) Corresponding author: Frederic Lens [E-mail: lens@nhn.leidenuniv.nl].

Figure 1. A simple ontology diagram showing a subset of wood anatomy terms from the *Plant Ontology*. This diagram illustrates the two most important relationships in the ontology (*is_a* and *part_of*) and shows how data annotations are associated with ontology terms. For example, a *tracheid* (PO:0000301) *is_a* type of *tracheary element* (PO:0000290) and all *tracheary elements* are *part_of* some *xylem* tissue (PO:0005352). Using the relationships specified in this ontology, a computer or human could infer that any *tracheid* is *part_of* some *xylem* tissue. Similarly, it can be inferred that a *tracheid* is a *plant cell* (PO:0009002), that *xylem* is a type of *plant tissue*, and that both are *plant structures* (PO:0009011).

The ontology also facilitates genomics studies through data annotations, such as genes expressed in plant structures represented by terms in the *Plant Ontology*. For example, *px1*, a class III peroxidase-encoding gene from *Picea abies*, is associated to *tracheid* via its effect on lignin biosynthesis during tracheid development (Marjamaa *et al*. 2006). Since a *tracheid* is *part_of xylem*, *xylem* also exhibits the phenotypic qualities of *px1*. Similarly, *PgMYB2,* a transcription factor that acts as regulator of lignin and phenylpropanoid metabolism during wood formation in *Picea glauca*, is associated to the general term *xylem,* since the experiment did not specify the exact cell type it was expressed in (Bedon *et al*. 2007). If a user searches the *Plant Ontology* for annotations on *xylem* and its subtypes, both *px1* and *PgMYB2* will be retrieved. If the homologs of these genes are known for other tree species, a researcher can query the mutant and gene expression data for those species to look for genes regulating wood quality and development.

February 5 to 7th, 2012, is a subset of the *Plant Ontology* with the goal of providing a uniform vocabulary and set of logical relations for wood anatomical characters, stages of wood development, and wood qualities. This wood-related vocabulary should be particularly useful for annotating molecular and genomic datasets for wood formation (*e.g.*, Schrader *et al.* 2004; Tuskan *et al.* 2006; Melzer *et al.* 2008; Dharmawardhana *et al.* 2010; Agusti *et al.* 2011) and establishing a semantic framework for comparative queries across data sets from various species. The *Wood Ontology* is thus being developed in collaboration with genomic resources for trees, such as *TreeGenes* (http:// dendrome.ucdavis.edu/treegenes) and the *Hardwood Genomics Project* (http://www. hardwoodgenomics.org).

In the *Plant Ontology*, the terms are each assigned a unique identifier (PO:xxxxxxx) and are connected by defined relationships. For example, the *is_a* relation specifies that one term (*e.g.*, a structure or development stage) is a subtype of another term. Similarly, the *part_of* relation describes how one term is a part of another term. For example, Figure 1 shows that every *tracheid* (PO:0000301) *is_a tracheary element* (PO:0000290), which *is_a plant cell* (PO:0009002), and every *tracheary element* is *part_of* some *xylem* tissue (PO:0005352), which *is_a vascular tissue* (PO:0009015). The network of terms in an ontology allows a user or a computer to interpret and analyze the relationships consistently. For example, using the relationships in Figure 1, a computer asked to retrieve data associated with the query term "*xylem*" would also return any data associated with *tracheary element*, and all of its subtypes, including, amongst others, *tracheid*.

Why is it important to develop such a resource for our IAWA community? Increasingly, wood scientists from different disciplines (anatomy, archeology, dendrology, genomics, mechanics, physiology, technology) are carrying out multidisciplinary research using terminology that may or may not be consistently applied. This often happens because two different communities either use similar terms to describe different entities or use different terms to describe the same entity, such as the words for juvenile, crown-formed, and core wood (Amarasekara & Denne 2002). The wood anatomical community has played a leading role in establishing a controlled glossary (IAWA Committee 1964) and lists of anatomical features useful for wood identification (IAWA Committee 1989, 2004) and bark identification (Trockenbrodt 1990). However, our efforts are an exception rather than the rule in the broader field of plant science, and the current glossary of terms is not interpreted in a relational context. By incorporating the IAWA glossary and feature lists into the PO, these vocabularies will not only be accessible to a wider audience, but will also be accessible to computer algorithms to perform complex database queries. For example, it would be possible to retrieve a list of genes involved in cell wall metabolism that are expressed in all of the cell types (or any given cell type) found in *xylem*. Thus, the *Wood Ontology* will be the means by which new genomic and molecular data are integrated with wood anatomical data. Finally, by expanding the PO to include a comprehensive set of terms associated with wood anatomy, we will forge new links between research on secondary growth and that of the larger plant biology community.

The *Wood Ontology* project will provide a structured vocabulary and database resource that will be valuable for all scientists, including the IAWA community. To maximize the utility of the resource and analyses it empowers, it is important for researchers to adopt the use of the ontology terms in the collection and dissemination of their data. The PO website (www.plantontology.org) is the main portal for the *Plant Ontology*, and presents the current version. Researchers can contribute to the *Wood Ontology* by suggesting terms or commenting on existing definitions using the *SourceForge tracker* site (http://sourceforge.net/tracker/?group_id=76834&atid=835555) after registering and logging in. Specific questions can also be sent via e-mail to the PO project by filling out the web based feedback form (http://plantontology.org/db/feedback/send_feedback?refer_to=/index.html), which can be reached via the "Feedback" link from the PO home page.

REFERENCES:

Agusti, J., R. Lichtenberger, M. Schwarz, L. Nehlin & T. Greb. 2011. Characterization of transcriptome remodeling during cambium formation identifies MOL1 and RUL1 as opposing regulators of secondary growth. PLoS Genet. 7: e1001312.

Amarasekara, H. & M.P. Denne. 2002. Effects of crown size on wood characteristics of Corsican pine in relation to definitions of juvenile wood, crown formed wood and core wood. Forestry 75: 51–61.

Ashburner, M., C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin & G. Sherlock. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. 25: 25–29.

Avraham, S., C.-W. Tung, K. Ilic, P. Jaiswal, E.A. Kellogg, S. McCouch, A. Pujar, L. Reiser, S.Y. Rhee, M.M. Sachs, M. Schaeffer, L. Stein, P. Stevens, L. Vincent, F. Zapata & D. Ware. 2008. The Plant Ontology Database: a community resource for plant structure and developmental stages controlled vocabulary and annotations. Nucleic Acids Res. 36, Database issue D449–D454.

Bedon, F., J. Grima-Pettenati & J. Mackay. 2007. Conifer R2R3-MYB transcription factors: sequence analyses and gene expression in wood-forming tissues of white spruce (*Picea glauca*). BMC Plant Biol. 7: 17.

Dharmawardhana, P., A.M. Brunner & S.H. Strauss. 2010. Genome-wide transcriptome analysis of the transition from primary to secondary stem development in *Populus trichocarpa*. BMC Genom. 11: 150.

IAWA Committee on Nomenclature. 1964. Multilingual glossary of terms used in wood anatomy. Konkordia, Winterthur, Switzerland. 185 pp.

IAWA Committee. 1989. IAWA list of microscopic features for hardwood identification. IAWA Bull. n.s. 10: 219–332.

IAWA Committee. 2004. IAWA list of microscopic features for softwood identification. IAWA J. 25: 1–70.

Jaiswal, P., S. Avraham, K. Ilic, E.A. Kellogg, S. McCouch, A. Pujar, L. Reiser, S.Y. Rhee, M.M. Sachs, M. Schaeffer, L. Stein, P. Stevens, L. Vincent, D. Ware & F. Zapata. 2005. Plant Ontology (PO): a controlled vocabulary of plant structures and growth stages. Comp. Funct. Genom. 6: 388–397.

Marjamaa, K., K. Hildén, E. Kukkola, M. Lehtonen, H. Holkeri, P. Haapaniemi, S. Koutaniemi, T.H. Teeri, K. Fagerstedt & T. Lundell. 2006. Cloning, characterization and localization of three novel class III peroxidases in lignifying xylem of Norway spruce (*Picea abies*). Plant Mol. Biol. 61: 719–732.

Melzer, S., F. Lens, J. Gennen, S. Vanneste, A. Rhode & T. Beeckman. 2008. Flowering-time genes modulate meristem determinacy and growth form in *Arabidopsis*. Nat. Genet. 40: 1489–1492.

Schrader, J., J. Nilsson, E. Mellerowicz, A. Berglund, P. Nilsson, M. Hertzberg & G. Sandberg. 2004. A high-resolution transcript profile across the wood-forming meristem of poplar identifies potential regulators of cambial stem cell identity. The Plant Cell 16: 2278–2292.

Smith, B., M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L.J. Goldberg, K. Eilbeck, A. Ireland, C.J. Mungall, The OBI Consortium, N. Leontis, P. Rocca-Serra, A. Ruttenberg, S.-A. Sanone, R.H. Scheuermann N. Shah, P.L. Whetzel & S. Lewis. 2007. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat. Biotech. 25: 1251–1255.

Trockenbrodt, M. 1990. Survey and discussion of the terminology used in bark anatomy. IAWA Bull. n.s. 11: 141–166.

Tuskan, G.A., S. DiFazio, S. Jansson, J. Bohlmann, I. Grigoriev, U. Hellsten, N. Putnam, S. Ralph, S. Rombauts, A. Salamov, J. Schein, L. Sterck, A. Aerts, R.R. Bhalerao, R.P. Bhalerao, D. Blaudez, W. Boerjan, A. Brun, A. Brunner, V. Busov, M. Campbell, J. Carlson, M. Chalot, J. Chapman, G.-L. Chen, D. Cooper, P.M. Coutinho, J. Couturier, S. Covert, Q. Cronk, R. Cunningham, J. Davis, S. Degroeve, A. Déjardin, C. dePamphilis, J. Detter, B. Dirks, I. Dubchak, S. Duplessis, J. Ehlting, B. Ellis, K. Gendler, D. Goodstein, M. Gribskov, J. Grimwood, A. Groover, L. Gunter, B. Hamberger, B. Heinze, Y. Helariutta, B. Henrissat, D. Holligan, R. Holt, W. Huang, N. Islam-Faridi, S. Jones, M. Jones-Rhoades, R. Jorgensen, C. Joshi, J. Kangasjärvi, J. Karlsson, C. Kelleher, R. Kirkpatrick, M. Kirst, A. Kohler, U. Kalluri, F. Larimer, J. Leebens-Mack, J.-C. Leplé, P. Locascio, Y. Lou, S. Lucas, F. Martin, B. Montanini, C. Napoli, D.R. Nelson, C. Nelson, K. Nieminen, O. Nilsson, V. Pereda, G. Peter, R. Philippe, G. Pilate, A. Poliakov, J. Razumovskaya, P. Richardson, C. Rinaldi, K. Ritland, P. Rouzé, D. Ryaboy, J. Schmutz, J. Schrader, B. Segerman, H. Shin, A. Siddiqui, F. Sterky, A. Terry, C.-J. Tsai, E. Uberbacher, P. Unneberg, J. Vahala, K. Wall, S. Wessler, G. Yang, T. Yin, C. Douglas, M. Marra, G. Sandberg, Y. Van de Peer & D. Rokhsar. 2006. The Genome of Black Cottonwood, *Populus trichocarpa* (Torr. & Gray). Science 313: 1596–1604.