RCN Plant Working Group notes, June 1-3 2011

Participants:
Eva Huala <huala@acoma.stanford.edu>
Peter Stevens <peter.stevens@mobot.org>
Pankaj Jaiswal <jaiswalp@science.oregonstate.edu>
Brent Mishler <bmishler@berkeley.edu>
Nico Cellinese <ncellinese@flmnh.ufl.edu>
Gavin Kennedy <gkam08@gmail.com>
Daniel Bunker <dbunker@njit.edu>
James Macklin <james.macklin@gmail.com>
Laurel Cooper <cooperl@science.oregonstate.edu>
Ramona Walls <rwalls@nybg.org>

Presentations and discussion:

Flora of North America project (James Macklin) http://www.efloras.org/flora_page.aspx?flora_id=1

- Semantic mark up of free text data
- Dynamic Flora, Eflora hosted at Missouri Botanic garden
controlled vocabs, parsing of descriptions, tying in specimens
-OCR does not work really well

Standards, conventions and controlled vocabularies in APWeb (Peter Stevens)
http://www.mobot.org/mobot/research/apweb/

Past vs Future:
comments on ontologies: Ontologies vs controlled vocabs
-best if small and well designed, "conventions", must be used
-monographic work- ignores the literature?, data mining relies on the literature - basic discrepancy
- taxonomy builds on 250+ years of work or does it? Info must be linked to individual specimens
-past literature used as a hypothesis to test (BM)
- need it to build up ontologies? Do we? Why?
PJ: need to go back and re-annotate the specimens- train the specialists to use the controlled vocabs
BM: software available to track samples etc: http://mooreabiocode.org/
Genius workflow: commercial product
-ovules shape is a quantity not a quality- continuous variation

Brent Mishler: Unified model for data for specimens, clades, taxa (Phylocode-style) and traits

instances: specimens and its associated measurements
traits: Intrinsic vs extrinsic-
extrinsic "traits": mega data, georeferences
inferred or directly attributed:

Introduction to RegNum (Nico Cellinese)

Phylogenetic definitions: Node, branch or apomorphy-based
crown clade: is a clade with all the terminals are extant or recently extinct
apomorphy-based= first instance of a character common to all
Apo-Angiospermae: 1st instance of a carpel
Regnum: PhyloCode name repository
http://www.ohio.edu/phylocode/
PhenoRegnum: search NCBO Bioportal and retrieve ontologies

<u>Introduction to Plant Ontology and Trait Ontology</u> (Pankaj Jaiswal)
(see general discussion notes below)

<u>High throughput phenotyping - ontology needs</u> (Gavin Kennedy) <u>http://www.plantphenomics.org.au/</u>

PlantScan -NIR images, moving away from LemnaTec platform
pseudo- 3D images
n-dimensional models
Challenges:
-translation from leaf to plant to canopy to landscape is difficult

Leaf Annotation:
- Leaf structure
- leaf topology- where is it located on plant?
- leaf numbering- record parameters on certain leaf at a certain spot
- Morphology, surface orientation in space
- leaf size, length and shape

Anatomy of the leaf: X-sections, mesophyll org, epidermis,
Stomata- location, distribution, #, size of aperture, guard cells

Leaf Function: Fluorescence, water content, chemical, gene expression, proteomics, metabolomics

Role of ontologies

<u>Ecological ontology development status and needs</u> (Dan Bunker)
Ecology and conservation
Scale from one site (BCI Panama) to globe, carbon storage,
Questions:
- seed size compare wild species and models, quantitative data
- id genes of importance, make predictions

TraitNet Ontology development
OBOE: Extensible Observation Ontology
observations, measurements, standards, precision, protocols
Pulling terms from TO, PO, GO, PFTO (?)
examples: seed mass
taxon phenotype annotations
Protege: OBOE plugin to add measurements of traits
has context: taxon, location, data collection, observer, genotype
nitrogen- part_of plant- BS would disagree?

<u>Brent Mishler:</u> Decompose Homology into Orthology and Paralogy:
Orthology- same structure in two different organisms, that arose from a common ancestor

[from wikipedia: Homologous sequences are orthologous if they were separated by a speciation event:
when a species diverges into two separate species, the divergent copies of a single gene in the resulting
species are said to be orthologous. Orthologs, or orthologous genes, are genes in different species that
are similar to each other because they originated by vertical descent from a single gene of the last
common ancestor.]

Paralogy:
[from wikipedia: Homologous sequences are paralogous if they were separated by a gene duplication
event: if a gene in an organism is duplicated to occupy two different positions in the same genome, then
the two copies are paralogous.
A set of sequences that are paralogous are called paralogs of each other. Paralogs typically have the
same or similar function, but sometimes do not: due to lack of the original selective pressure upon one
copy of the duplicated gene, this copy is free to mutate and acquire new functions.]

relationships- is_orthologous_to
leaf in pine is_orthologous_to leaf in magnolia, but not a leaf in moss,

relationships-
Leaf in moss is_not_homologous_to leaf in magnolia

Plant Working Group general discussion topics:
1. Discuss whether PO (Plant Ontology) can serve as the reference plant anatomy ontology for systematists as well as model organism biologists and come up with a list of specific changes that will be needed.

2. Explore how links can be made between PO and taxonomic databases to define the taxonomic scope of terms in PO.

3. Clarify the role of TO (trait ontology) - should it be used as a set of pre-composed EQ terms for all plant taxa and all areas of plant biology including systematics, ecology and model organism biology?

General discussion notes:

child of phyllome:
Phyllome: definition discussion:
A lateral plant organ produced by a shoot apical meristerm - needs work
Link to SourceForge tracker: http://sourceforge.net/tracker/?group_id=76834&atid=835555

Non-vascular and vascular leaf issue- needs to make it easier to annotate, is term inflation an issue?

BM: mosses-hydroids and leptoids are vascular tissue? conducting tissue is a better overarching name

-non-vascular leaf- definition needs work: Leaf that is part of a gametophore?

cigar leaf: phenotype/ development phase? not a intersection term lacking term from PATO
- consider moving to TO?
- annotation to post-composed terms?

Peter: Do we have a class for caulome? RW: maybe could be synonym for shoot axis.

Should the PO be the standard for definitions for all plant scientists? Should we get rid of other glossaries like APWeb? It should be an authority in the sense that it is based on community agreement, not in the sense that it is telling people what they must do.

We can add alternate definitions -- easy in OWL. These can match different definitions to different labels. But definitions would have to make sure that the definitions mean the same thing. This could be complementary to what is in the comments.   Example: definitions that are worded in a more user friendly style than the standard genus-differentia definition.

Melissa Haendel: Uberon Ontology;
- keep the homology separate, don't need to reason across it here, keep in a table with attributions, with explicit statements
-See Homology Ontology in bioportal- also Phenoscape group

http://bioportal.bioontology.org/ontologies/1328/?p=terms#details

See: Roux, J. and Robinson-Rechavi, M. (2010). An ontology to clarify homology-related concepts. Trends in Genetics, 26, 99-102.

Suggestions for PO:
1. Provide a web page that is the PO terms listed as a glossary (alphabetical). Must update whenever file updates. Must link back to PO page for each term.
2. Work with PS to create and include the terms from his glossary, will eventually be retired. Use definitions that are worded in a more user-friendly style than the standard genus-differentia definitions.

3. jQuery widget for search field for autofill.

4. Add a place for commentary on the PO website.

5. Link person's name to the definitions that they work on

6. *Do more outreach: Maybe host workshops at Botany meetings on Sunday before sessions (NC)

Eva: Can the PO serve as a reference ontology for plants?
BM: conditionally yes, focus on structure,
separate function, development homology into another place
-allow cross-classification among these factors

NC: community effort, involvement and buy in, workshops at meetings, use cases, demo utility

EH: involve domain experts, outreach efforts,

MH: discussion link on web page?

JM: collaborative work, build use cases, reasoning and fostering research,
needs: how deep does the PO want to go on shapes etc, standardized references, date stamps

DB: PO is an excellent research, works well for TraitNet, not sure if the TO is adequate currently, precomposed terms, need to search on traits, observations and qualities , quantify traits

PS: can't tell people what to do, simplify botanical terms, needs better use cases, glossary would make it more user-friendly, links from AP web,

GK: ditto DB!, Examples: leaf positioning, leaf orientation, both traits captured at a point in time, can express the temporal aspect, interested to engage with PO, integrate with POOD
PO and TO qualitative terms,

Wrap-up:
Achievements:
- We agreed (more or less) that PO is the central reference ontology for plants, the TO
- Developed new collaborations
- Improved communication between the different groups- ecology, systematics, taxonomy HPT phenotyping,

Goals:
- character atlas or matrix focused on leaf, between FNA and sources from books etc.
PJ: Add new class to PO? annotation tool,

- Use case- how will this be developed? Must be compelling, could focus on leaves?
PJ: HALS demo leaf development genes
examples of mutations -peltate leaves

- Improve outreach- get workshops out to more groups eg. Botany meetings

- images associated with the terms, layers

Post-meeting email discussion notes:

From Eva's email (6-7-11):
The two main plant working group goals (LC: and one smaller goal) we laid out at the end the workshop were:

1. **a. Collect an initial set of leaf-related characters from FNA (James Macklin) and the textbook: Manual of Leaf Architecture (Ellis et al, 2009),** suggested by Ramona Walls.

EH: If I understood correctly this could be achieved by the end of the summer

1. **b. The set of leaf characters could be mapped to existing and newly requested PO and PATO terms** (although we didn't discuss how to do this or who could do it).

JM: I like the idea of having two focused goals. As I indicated, I should be able to get a list of terms based on Flora of North America (FNA) to the PO and PATO folks sometime in early Fall once we have the database completely populated.

RW: I am quite interested in both of the goals we discussed, as well as the homology goals that we discussed at the break-out meeting. My primary obligation is to fulfill the PO grant objectives, but I am very happy to work on these goals as time allows. I have made preliminary inquiries into using the images and definitions from the "Manual of Leaf Architecture" and it seems likely that we will be able to. I am waiting for more formal confirmation, but for the time being, we can at least use it as a source for terms that we will need.

GK: For goal 1, can you clarify these 'characters' for me? My understanding from the meeting was that we would be seeking to define a set of EQ terms around the leaf, but now I'm not so sure. If so, then we are definitely happy to contribute there, once I meet up with my biologists.

2. **Develop a compelling use case/scientific question that illustrates the usefulness of the ontology approach.**

The use case would need to make use of PO's ontology structure and span model organism biology and systematics.

Options for the Use Case(s):
- The genetic basis for evolution of peltate leaves, combining data on mutations causing peltate leaves in non-peltate species and data on plant species with peltate leaves.

- Explore the genetic/ evolutionary basis for simple vs. compound leaves*

*EH: I'm thinking there may be more data available for this option and we may be able to make better use of the ontology structure because it's a broader category.

Responses:
NC: I'll be very happy to contribute to goal 2. I think such a study is clearly missing and a paper similar to Washington et al. 2009 addressing plant examples would be fantastic. There has been quite a bit of work with flower development too.

BM: I'd be glad to pitch into the second goal. It might be a good idea to have two use-cases, one related to evo-devo as in the peltate leaf idea, the other related to ecology-phylogeny, e.g., SLA in relation to climate variables. This would let us reach out to two different audiences to show the utility of ontologies.

BM: One other thing we said we would do is work on a homology ontology using leaves as an example that would cut across the structure-based PO terms. I heard more discussion happened at the Phenoscape meeting right after ours, along these lines. I think we plant folks should stay involved in that discussion.

RW: I am quite interested in both of the goals we discussed, as well as the homology goals

DB: I'd be happy to contribute to goal 2.
Building on Brent's suggestion to develop two UCSs, it would be very powerful if we can develop a single use case scenario that connects genes and model orgs all the way to wild organisms (specimen collections?), via phenotypes in the middle. That way we can use phenotypes to bridge the gap and use data from one realm to benefit the other.
Looking forward to developing the UCSs further.

GK: For goal 2, my use case expectations would be fulfilled if, as Dan suggested, we can connect genes to phenotypes across model organisms, crop species and up to wild species. I don't think we discussed this much, but the ability to introduce environmental factors into the use case(s) would be good, e.g. anatomical entity + phenotypic trait is a consequence of gene/genome variation X environment/environmental treatment. e.g. stomata + conductance influenced by gene:OST1 + soil moisture content. This is my own musing so I'm sure more sophisticated examples can be produced.

3. There was a request for PO to provide access to the Plant Ontology anatomy terms in the form of an alphabetized, browse-able glossary-like list of terms and definitions.

This will make it easier for people not familiar with ontologies to make use of the PO terms and could also provide a centralized forum to discuss the terms and their definitions.

Responses:
PS: I will be in contact with the PO group to make this comprehensive list of terms – a controlled vocabulary - available, adding Apweb and other terms (inc. any leaf terms, etc.). We may need a subgroup for this, but let's see how it goes.

JM: I will coordinate with Peter so we can try to figure out how to merge terms and definitions. The PO team can then help us figure out the best way to get the terms loaded in.

EH - To push these forward, we could organize into two separate groups to push forward the two main goals above.
Eventually these two threads would come together again when we're ready to try to answer the scientific question from goal 2 using leaf characters associated to taxonomic groups from goal 1, plus mutant phenotype and gene expression data from other sources.

From PJ: We (the PO) have started discussing our next goals, both short and long term. This includes leaf character matrices and glossary style PO presentation on the web

LC: I am interested in contributing to the efforts of each goal. We will keep in contact with the group as we move forward on various fronts. Initially, it seems we need to clarify the goals, rank them in terms of importance, set up the interest groups and set up a timeline.