Phenology Collaboration

Virtual Meeting on 27.11.2017

Link to the online meeting :

https://hangouts.google.com/call/dUXK3tQGa4r5PPxwDz1vAAEI

Schedule :
Costa Rica : 9.30 am
US, California : 7.30 am
US, Florida : 10.30 am
US, Connecticut : 10.30 am
Canada, Ottawa : 10.30 am
France : 16.30 pm

Participants :
Costa Rica (TEC Univ.) : José Carranza.
California (San Luis Obispo, Cal Poly) : Jenn Yost.
California (iDigBio) : Libby Ellwood.
Florida (iDigBio, Univ. of Florida) : Pamela Soltis, Gil Nelson, Katie Pearson (Florida State University).
Connecticut (New Haven, Yale Univ.) : Patrick Sweeney.
Canada (Ottawa, Agriculture and Agri-Food Canada) : Joel Sachs.
France : Alexis Joly (Inria), Hervé Goëau (Cirad), Erick Mata Montero (TEC Univ.), Pierre Bonnet (Cirad).

Points of discussion :

. Feedback on recent research results obtained on evaluation of [Machine learning](#) for automated species identification based on herbarium data. (link with bird identification from spectrograms ; link between plant identification & birds migration : distinction between train / test set).

. Exchanges on visual data analysis for automated phenology annotation, from French and Costa Rican participants (discussion on controlled vocabulary used by Katie ; NEVP standard phenology ontology ; Shared paper on the Subject ; Potential Train datasets (Patrick)).

. What are the most needed information for herbarium users, (i) at large and (ii) small taxonomic sale? (ex. : Search on iDigBio portal for specimens of *Fabaceae* in flower only ; Search for specimens on *Convolvulus* in fruit and flower only ; Search for specimens of *Asteraceae* with flowering buds ; etc.).

. Available annotated datasets / Data availability :
- Tree species of French Guyana : http://publish.plantnet-project.org/project/caypub_en
- iDigBio dataset annotated by Katelin Pearson, Asteraceae, ~20,000 specimens available as of November 2017 (exchanges with Katie and Libby)
- Potential central repository for data suggested by Joel.
- Discussion on the potential frameworks used for automated identification.

. Standards related to phenological states annotation on herbarium.

. Potential experiments of automated annotation of herbarium specimens :
- Large tropical family with an high number of tree species : *Fabaceae, Rubiaceae, Myrtaceae, Euphorbiaceae* (-> CAY dataset). But also potentially *Rosaceae, Sapindaceae, Betulaceae & Fagaceae, Ericaceae* (-). Potential involvement of students from UF, FSU, and Yale, potentially using crowdsourcing platforms such as "[Notes from Nature](#)" and images from iDigBio
- Key species and/or genera of interest for scientists who have small annotated datasets
- [GN] perhaps use specimens from the FSU herbarium that have been scored for phenology as a test dataset.

. [GN] Consider a workshop proposal to NSF for funding to get this and a larger group together focused on phenology and deep learning; potentially follow the

workshop or simultaneously submit an NSF RCN proposal that can provide funding for some of our work.

[PS] Following workshop and/or RCN funding, consider a proposal to NSF's Advances in Biological Informatics - Development Option to develop 'industrial strength' platform using web services, etc. to provide resource for the broader community (perhaps Sept 2019). Investigate possible international funding options.

. Potential calls for collaboration.

Meeting #2: February 8, 2018
- We need at least 100-200 annotated specimens per family (ideally)
- We will add the French Guiana, NEVP, and FSU data to produce these datasets
- Patrick, Katie, and anyone else with access to data will get counts on how many annotated specimens we currently have available (imaged) and will communicate with Pierre et al. about how to access those data and images
- Datasets put together by Anna Monfils?
- The minimum data we need is URL, species name, specimen number/uuid, phenological annotations
- Pierre et al. will send a list of what they need, we will look at our datasets and send the data that we have
- Goal: be able to join our data by the end of March

Notes on dataset interoperability:
- The French Guiana dataset is coded as reproductive? yes/no (not including buds)
- The NEVP dataset is a hierarchical coding scheme reproductive? → which stages?
- FSU data is buds/flowers/fruits? true/false
- NEVP and FSU data can easily be transformed into reproductive? yes/no
- Katie's Asteraceae dataset is buds/flowers/fruits? true/false

Important info : Url, species, specimen number, pheno. annotation.

Family	Number of annotated/imaged French Guiana specimens	Number of annotated/imaged NEVP specimens	Number of annotated/imaged FSU specimens
Acanthaceae	710	25	400
Anacardiaceae	343	229	241

Annonaceae	2780	7	46
Apocynaceae	2416	691	183
Araceae	1711	487	93
Arecaceae	3114	5	78
Asteraceae	1790	5550	9755+
Bignoniaceae	1540	7	29
Bromeliaceae	1187	8	65
Burseraceae	1027	0	0
Rosaceae	0	16150	1107
Fabaceae	3270	4433	8146
Malvaceae	288	1092	486
Ericaceae	258	13283	1688