Group 11

Myla Aronson, Mason Heberling, Alex White, Hervé Goëau, Titouan Lorieul, Renato Figueiredo, Nicky Nicolson, Nadya Williams, Annika Smith, Emily Meineke

Beyond this workshop, how might we develop a tool that could be easily adapted to other applications (insect damage, morphometrics, leaf shape/size)?

Key question
Can we translate methods for reproductive phenology scoring with ML to other traits (e.g., herbivory damage) - transfer learning from phenology-based model to leaf trait model?

A: YES! Classification models could be used to identify specimens with damage, then go finer scale from there. Use to mask images to produce image of leaves for analysis

Potential challenges
One key consideration for what traits might be able to be tracked is how big the training datasets will be. For some things (e.g., leaf size), training data are widely available on specimens, while specific types of herbivory might be rarer? A related question is for the appropriate algorithms, how big do training datasets need to be?
A: This technology of image recognition is rapidly improving, which may help with the training data issue, i.e., maybe less training data will be required than we expect now.

Challenge is how the plants are pressed, overlapping structures. Can this be scored itself, e.g., automated via citizen science? One thing to keep in mind is to create training datasets on ‘messy’ specimens, because restricting to beautiful specimens might make for a less accurate model.

Segmentation will require annotated images for training (e.g. outlining leaves w LabelMe) -- Segmentation is identifying what individual pixels are, e.g., leaf, herbarium sheet, flower. Develop a workflow for annotating images-- what platforms are available?

-annotations tied to terms in Plant Ontology?

Hierarchy of annotating image-- generalize

What are the collector biases in these additional traits? They have to be considered when developing hypotheses that can reasonably be tested and approaches.

How do we build cyberinfrastructure and workflows that can serve multiple machine learning applications on digitized specimens? (Can we have a hierarchy of outputs that allow users to make use of multiple annotations on specimens (e.g, plant vs. sheet, then leaf, flower, stem)?

Major areas of interest for applying machine learning beyond phenology
How plant traits, other than phenology, change over time and across the plant phylogeny? -Leaf size, flower size
How insect damage has changed over time?
Utilize herbarium samples to examine leaf/floral evolution within clades. (what data are required for this?)

Mason's awesome table of traits that can be extracted from specimens:
Table. Phenotypic traits that can be measured using herbarium specimens. *Functional traits that have been used to address global change hypotheses and rely on or directly benefit from digitized specimen images are also indicated. See Supplementary Materials (table S3) for an extended version with references.*

Trait (units)	Biological relevance	Global change context?	Use images?
Organismal-level			
Plant height (mm) Internode length (mm)	Plant strategy axis: competitive ability, stress tolerance; response to abiotic/biotic stressors	X	X
Pathogen/symbiont occurrence (presence/absence)	Disease dynamics and susceptibility; associations with endophytes, mycorrhizal fungi, and other symbionts		X
Leaf-level			
Leaf size, area (mm^2)	Light interception, water balance, nutrient status	X	X
Leaf mass per area (mm^2 g^{-1})	Resource economics: carbon investment, photosynthetic rate, leaf lifespan, construction cost	X	
Vein density ($mm\ mm^{-2}$)	Carbon economics: hydraulic and photosynthetic capacity	X	X
Leaf thickness (mm)	Structural investments		
Tissue chemistry			

Nutrients (e.g., % N, C, P, S)	Nutrient strategy and/or nutrient availability		
Heavy metals (% Pb, Zn, Cu)	Heavy metal accumulation; stable tracers of environmental pollution	X	
Stable isotopes (e.g., $\Delta^{13}C$, $\delta^{13}C$, $\delta^{15}N$, $\delta^{18}O$)	Water-use efficiency; past environmental conditions	X	
Deuterium isotopomer ratios	Proxy for long-term shifts in carbon metabolism	X	
Macromolecules (%)	Protein, fatty acid, and amino acid composition	X	
Secondary metabolites (e.g., % flavone)	Defense against herbivory or UV damage	X	
Leaf shape (e.g., dissection, toothiness, width:length; geometric morphometrics)	Thermoregulation, water balance	X	X
Trichome density (hairs/cm^2) Trichome length (cm^2)	Herbivore defense, water conservation, UV-B protection		X
Stomata related traits (e.g., density [no./mm^2])	Water-use efficiency, gas exchange	X	X
Herbivory damage (%)	Insect damage	X	X
Flower-level			
Seed chemistry (e.g., µg foranocoumarins)	Secondary metabolites, lipid composition	X	
Flower/fruit morphometrics (e.g., size, number, seed number per fruit, seed mass)	Reproductive output, plant-pollinator interactions, dispersal distance/colonization ability, seed output, seedling survival	X	X
Spore size	Reproductive biology, ploidy level, dispersal ability		
Ultraviolet floral pigmentation	Plant-pollinator interactions		
Pollen identity, size, chemistry, protein content	Pollen transport network ecology; food quality for pollinators	X	
Phenology (phenophase)			

Leaf-out date	Growing season length, understory light acquisition, responsiveness to environmental cues and changes	X	X
Fruit maturation (e.g., first fruit, peak fruiting)	Seed dispersal, senescence in response to environment	X	X
Flowering time (e.g., first flower, peak flowering)	Reproductive opportunity, plant-pollinator interactions, responsiveness to environmental cues and changes	X	X