

Deliverable		n° D6.1	
Deliverable title: Simplified model of Arabidopsis thaliana			
Task	WP6 - Plant modelling T6.1 Modelling		
Task Leader	Inria	Planned Date	31.10.2018
		Effective Date	01.11.2018

	Written by	Reviewed and approved by		Authorized by
Name / Surname	Ch. Godin	F. Besnard	P. Hanappe	
Date	October 2018			

1 Executive Summary	2
1.1 Summary of deliverable content and initial objectives	2
1.2 Partners involved	2
1.3 Relation with other work packages and tasks	2
1.4 Links to videos, flyers, ...	3
1.5 Dissemination / IPR policy	3
2 Simplified model of Arabidopsis thaliana	4
2.1 First approach: simple geometric model of Arabidopsis inflorescence stem	4
2.2 Second approach: Realistic model of Arabidopsis thaliana architecture development	5
2.3 Estimation of sequence of angles from virtual point clouds (interaction WP6-WP5)	9
2.4 Work on the simulation platform	11
3 First investigations for D6.4: Results of trained models	12
4 Bibliography	13

1 Executive Summary

1.1 Summary of deliverable content and initial objectives

This report describes the work carried out in WP6 through a description of the deliverables. Only D6.1 was due to this first-year milestone (design of a simplified model of *Arabidopsis thaliana*). However, since we also started to work on D6.4 (training algorithms), we report on this as well at the end of the text.

WP6 is dedicated to model plant architectures. Such models are used in ROMI for two main purposes.

First, they can help assessing 3D-reconstruction algorithms by providing a known synthetic ground truth through simulated synthetic data. 3D-reconstruction algorithms are developed in ROMI WP5 and are described in deliverable D5.3, available for this first review. The overall processing pipeline tries to extract precise architectural parameters (for instance phyllotaxis), by imaging a plant in 3D. It involves a 3D-reconstruction of the plant from RGB images, its segmentation into the main stem and lateral elements (branch, leaf or flower) and finally the measurements of interest (i.e. divergence angles between organs and lengths of internodes). Assessing the absolute accuracy and precision of such pipeline is impossible with real plants, since the technical uncertainty cannot be distinguished from the inevitable biological variability. In this report, D6.1 precisely describes how we generated plant architecture models to evaluate the performance of image analysis pipelines involving 3D-reconstruction.

Second, models can guide reconstruction algorithms operating on point clouds obtained from images by providing a prior knowledge of the expected architecture. In particular, machine learning techniques could use such models to infer relevant plant architecture parameters from real images after a training with models. In this report, the part D6.4 describes the first investigations in this direction.

1.2 Partners involved

Leader: **Inria**
Participants: **CNRS**

1.3 Relation with other work packages and tasks

Relation to WP5: WP6 provides support for the model-based 3D-reconstruction algorithms. The plant models are also used to generate synthetic images to train neural networks in WP5.

1.4 Links to videos, flyers, ...

A movie showing the dynamics of growth of the plant is available here:

https://drive.google.com/open?id=1_2pgXwfj8jV8WGUmDM6CZgoREU4grvwL

1.5 Dissemination / IPR policy

The code of the Arabidopsis thaliana plant model is available at **TODO**.

2 Simplified model of *Arabidopsis thaliana*

During this first year, we developed two methods for generating plant architectures. The first one is a simple procedural method specific to the targeted application, i.e. extracting angles between consecutive lateral organs on the main stem of the *Arabidopsis* inflorescence described in D5.3. This model was instrumental in getting a first rapid virtual geometric model of the plant stem on which to test our pipeline for extracting angles. The second is a more generic procedural method where the development of the plant architecture is modeled in its entire complexity at organ resolution. This developmental model, whose design is more complex, can be used in a wide range of application contexts, including detection of angles between organs as in the first simple model, but also to address any arbitrary questions related to the reconstruction and analysis of plant phenotypes.

2.1 First approach: simple geometric model of *Arabidopsis* inflorescence stem

In this approach, a static geometric model of the *Arabidopsis* inflorescence stem is built. The main shoot is modeled as a cylinder. A thin cylinder is used to model each peduncle and a slightly larger and longer cylinder is representing the organs attached to it. The organ and its peduncle are then attached along the main shoot. The parameters describing the model are:

- the length of the main shoot, peduncle and organ lengths and radii ($L_s, L_p, L_f, R_s, R_p, R_f$).
- the angle of organs relative to the main shoot (α_s).
- the internode l , that is the distance between two consecutive attachment points of the organs.
- the relative angle in the plane orthogonal to the main shoot between to consecutive organs (α_i)

Angles α_s and α_i are the same for all organs. The internode distance is modeled as a function of the distance of the attachment point from the main shoot terminal $l(z) = A/(1+\exp(-a.(z-b)))$ in a similar fashion to what is reported about real plants (Mundermann, 2005).

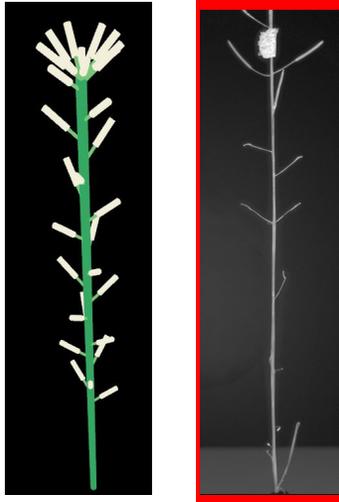


Figure 1: Inflorescence stem of *Arabidopsis thaliana* showing lateral siliques generated by a simple geometric model and a picture of a real plant inflorescence stem

2.2 Second approach: Realistic model of *Arabidopsis thaliana* architecture development

This second approach is more generic: the plant architecture is fully represented and results from a developmental model. In this first year, we designed a model of *Arabidopsis thaliana* development.

The idea consists in modeling the activity of apical meristems that build-up all aerial plant organs. Apical meristems are undifferentiated population of cells at the tip of each plant stem and that produce internodes and lateral organs (leaves, flowers, etc.) as they grow. When created by meristems, these organs have a very tiny volume and are hardly visible to the naked eye. However, they then progressively elongate up to a limit volume and size where they stop elongating. The apical meristem also produces lateral meristems at the basis of lateral leaf that can enter into activity in turn with a possible delay. This ensures a form of recursivity in plant growth that establishes other axes and produces the plant branching system. The laws of organ production and elongation have been reported in the literature for different species and growth conditions. Here we used a mixture of knowledge coming from literature on *Arabidopsis* development (e.g. Mundermann et al., 2005, Besnard et al. 2014).

To model rules describing meristem activity, we used the formalism of L-systems (Prusinkiewicz et al., 1990, Godin et al. 2005). In this formalism, a plant stem is represented by a string of elementary modules. Consecutive modules in the string correspond to consecutive organs (internode, leaf, flower, meristem) in the stem. Strings can be nested using parentheses to represent plant branching networks of stems.

Starting from a string, called the “axiom”, and representing the initial state of the plant branching system, the modeller defines a number of development rules (called rewriting rules) that express how

each type of organ changes throughout time and within its context in the plant (i.e. depending on its local neighbors). Rewriting rules related to meristem productions correspond to plant organogenesis. Rewriting rules related to change of other organs (leaf, internode, flowers, flower organs) express how the shape, volume and orientation of these organs changes according to time and context.

In the case of *Arabidopsis*, we assumed that meristems have a state proceeding from vegetative (immediately after germination) to mixed (vegetative-inflorescence), and then to inflorescence. In vegetative state, meristems produce leaves and lateral meristems that stay dormant. Internodes do not elongate. In mixed state, meristem produce internodes that elongate, leaves and lateral meristem that can enter themselves into growth after a delay. Finally, in floral state, meristem only produce lateral flowers, that later turn into fruits (siliques). The meristem stays a number of steps in each state that is determined by an order 1 Markovian model, with transition probabilities P_{ij} and initial probabilities P_i such that $p_0 = 1$ and $p_1 = p_2 = 0$ meaning that plant necessarily start in a vegetative state (i.e. by producing the leaf rosette). We also allowed in this first version the production of lateral meristem only by meristems of the main axis, thus limiting branching order in the plant to 2 (1 = main stem, 2 = axillary branches).

The angle between consecutive organs is assumed to be drawn from a gaussian distribution centered on the golden angle $\phi = 137.5$, with a standard deviation σ_0 that can be controlled as a model parameter. Similarly, the initial insertion angle of branches and organs onto their bearing internode at the moment of creation is controlled by a normal distribution $N(\alpha, \sigma_1)$. This angle opens with time up to a limit angle, at a speed v_1 . This mimics the augmentation of insertion angles generally observed on plant as a function of the distance to the bearing axis' tip. Likewise, a growth rule describes the elongation of internodes from a value close to zero at internode initiation to a value drawn in a normal distribution $N(\lambda, \sigma_2)$. This growth is assumed to be sigmoidal in time, with parameter that can be controlled by the modeller.

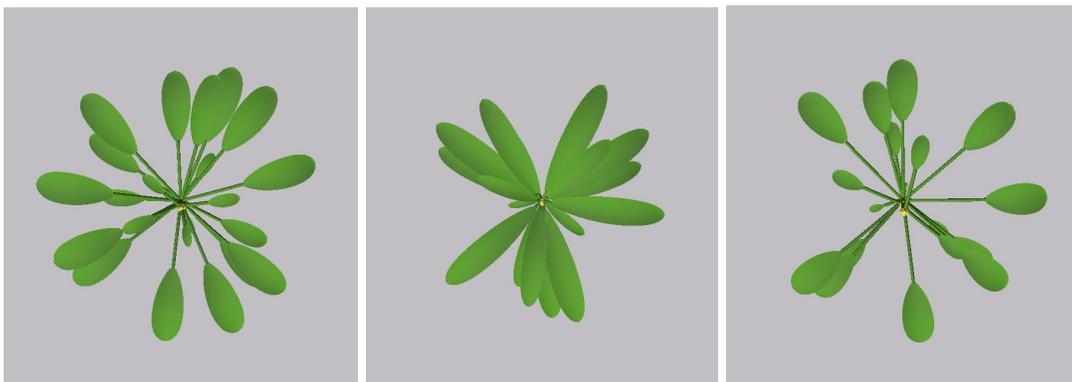


Figure 2: Variations in the shape of rosette leaves corresponding to contrasted phenotypes.

In real plants the geometry of axes and other organs such as leaves, flower petals or siliques, is highly dependent on phenotypes. To accommodate for such a plasticity, we added the possibility to select organ shapes in a probabilistic manner. The geometry of leaves for instance, composed of a petiole and a leaf blade, is controlled by a size, an allometric ratio (length/width), a proportion factor between the petiole length and the leaf blade length. In this way it is possible to reproduce a large variability of leaf shapes observed in plant phenotypes (Weight et al. 2008, Bensmihen et al. 2008), Figure 2. At a more global scale, the global geometry of axes is not straight as it is affected by gravitropism (axes tend to orient straight up against gravity). This factor can be controlled in the model parameters.

Based on these growth rules and stochastic parameters, various phenotypes of 3D plant architectures can be simulated by mimicking in a realistic manner the plant growth process at the level of organs (Figure 3). A movie showing the dynamics of growth of the plant is available here: https://drive.google.com/open?id=1_2pgXwFj8jV8WGUmDM6CZgoREU4grvWL.



Figure 3: Various runs of the *Arabidopsis* simulation model showing plant architecture variability due to model stochasticity.

These 3D virtual plants can be downgraded to simulate the architecture of the main inflorescence stem by deactivating the possibility of branching in the mixed vegetative-inflorescence state (Figure 4).



Figure 4: Generation of series of *Arabidopsis* inflorescence stems using our stochastic model where branching in mixed state is deactivated.

These stems are then used in our pipeline as input synthetic data on which we simulate 2D RGB photographs around the stem on a circular trajectory centered on the plant at regular angles (see next paragraph).

For plant architecture models, our next step in ROMI/WP6 will be threefold:

- make use of the variability of these virtual plants to train machine learning systems to reconstruct 3D architectures from a series of images;
- develop developmental models of other crop and weed species;
- design, develop and test a new type of plant architecture model, not based on developmental process, to be used as a prior knowledge in the reconstruction of plant structures from 3D point clouds.

2.3 Estimation of sequence of angles from virtual point clouds (interaction WP6-WP5)

We started to use our plant architecture models in interaction with plant architecture inference pipeline carried out in WP5.

For the simple geometric model, views of the model were rendered along a circular trajectory around the plant using the Open3D interface to OpenGL (Zhou, 2018). For the LPy model of plant architecture we used a similar approach with built-in algorithms available in the PlanGL library (used by LPy for 3D geometric modeling). The developmental program first simulates a plant architecture and then proceeds to the simulation of a series of circular photographs at the end of the simulation. All simulated photos are recorded on the disk for future use within the reconstruction pipeline (WP5).



Figure 5: Circular photos taken by a virtual camera (here 4 angles are shown at respectively 0, 90, 180, 270 degrees).

These simulated pictures made it possible to test the entire pipeline, from synthetic data generation to angle detection based on image analysis and to compare the reconstruction angles with the ground truth coming from the synthetic data.

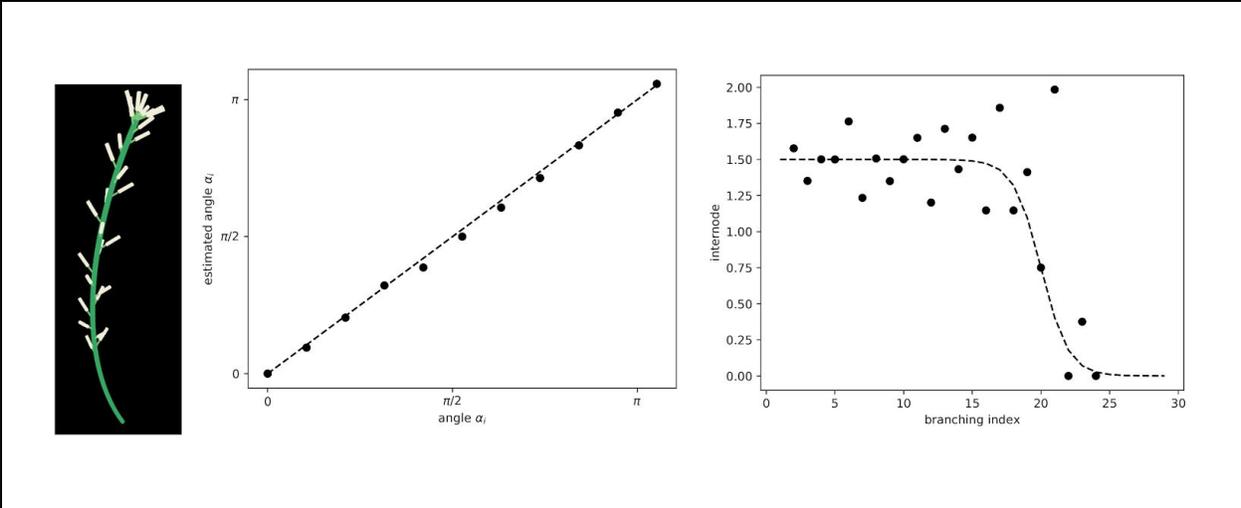


Figure 6: (Left) Rendered view of a model *Arabidopsis*. (Center) Estimation of angle between 2 consecutive fruits α_i based on 100 views and voxel size of 2mm. (Right) Sequence of internodes as measured (dots) compared to the internodes imposed at generation (dashed line).

The angle estimation algorithm was evaluated on pairs of organs with various angles. Figure 6 present the result of the method, and compares estimated angles and estimated internode length to the ground truth in the simulated experiments.

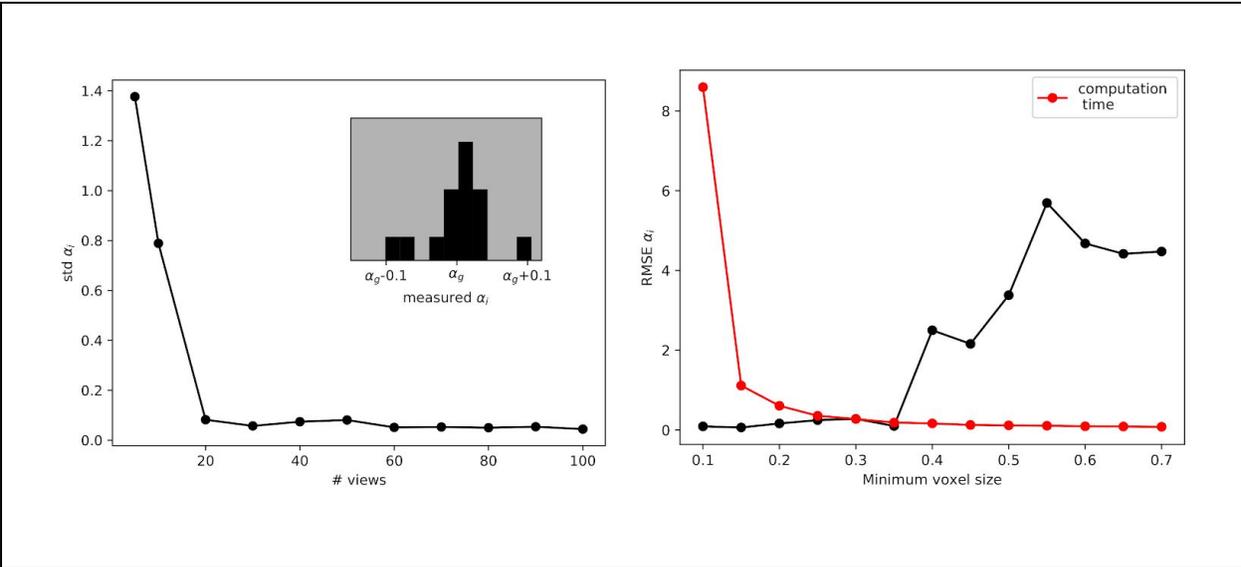


Figure 7: (Left) Standard deviation of the measured angles depending on the number of views considered to make the estimation. The insert shows a sample histogram when considering 100 views. (Right) RMSE of the estimated angles depending on the voxel size. We also report the computation time depending on the voxel size (arbitrary units) in red.

In Figure 7, we present the variance of angle estimates as a function of both the number of views and the minimum voxel size. We can see that in each case, there is a threshold value where the reconstruction fails. Estimation of the angles are good when more than 20 view are taken around the object. Figure 2 also shows that voxel size should be kept low enough to have a small root mean square error in estimation. When voxel size is too small, the computation time explodes and there is thus an optimal window of voxel sizes (between 2mm and 3.5mm) for which a good estimation may be obtained at reasonable computation cost. Variance around the golden angle is observed as around 20 degrees (Guédon et al., 2013) and the accuracy achieved by our method of less than 5 should therefore be largely sufficient to observe anomalies in the angle sequences.

All programmes developed in WP6 are available for the ROMI partners and will be made available under open-software licence after publication.

2.4 Work on the simulation platform

We also worked during this first year on the re-design of our simulation platform for plant growth modeling. This will make it possible to develop a new type of plant architecture simulation framework, more adapted to the interpretation of image or scanner data.

Our simulation software to model plant development is LPy, a python-based L-system language (Boudon et al, 2012), Figure 8. Currently LPy is available through the openalea platform (Pradal et al., 2008). Before the project started, it was decided to re-design the openalea platform, based on a new software engineering approach, so that the system can scale up more easily with the applications. Such a work will impact deeply ROMI as the new system will be used for the development of the algorithms used for reconstructing plant architectures from image data. After one year, a prototype of the new platform, Gnomon, is already available. At the beginning of next year, we will port the LPy and MTG libraries to the new system. This will make it possible to define ROMI pipelines in a user-friendly way and share them within the consortium.

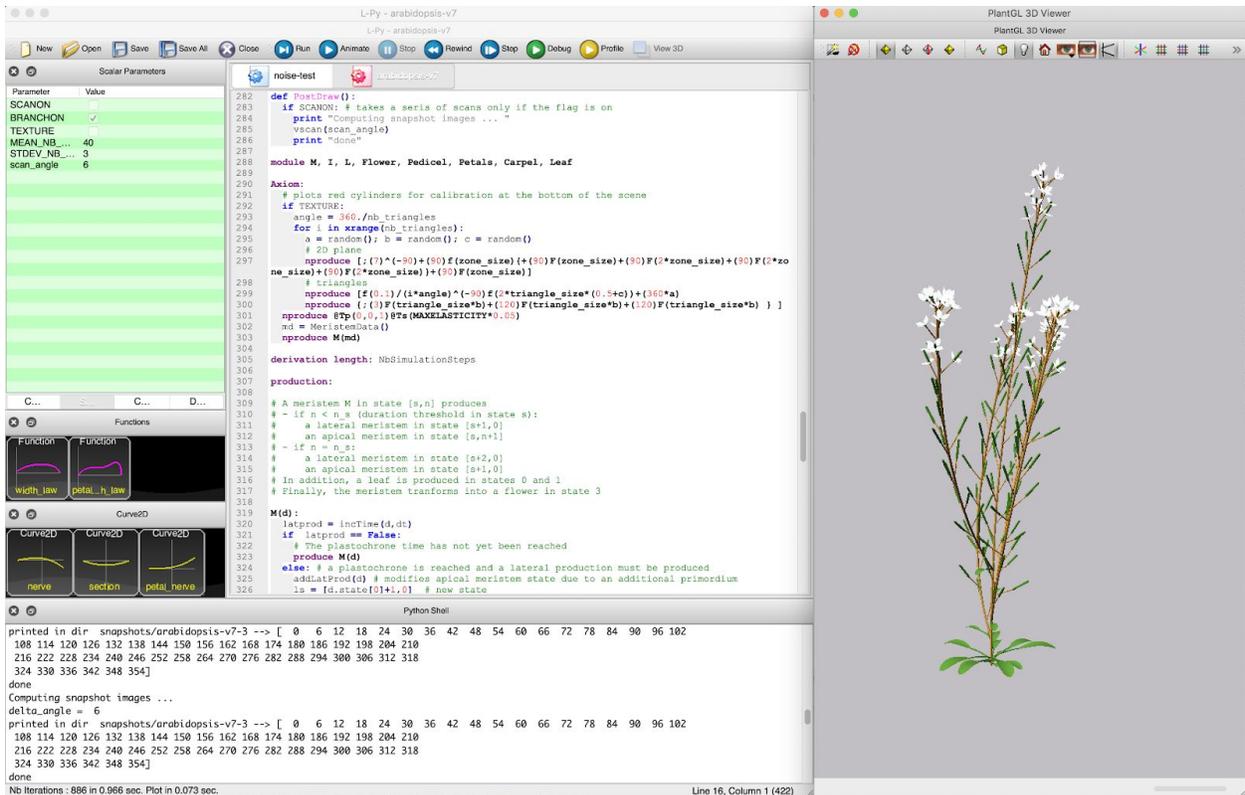


Figure 8: LPy environment for modeling plant growth.

3 First investigations for D6.4: Results of trained models

We anticipated the work in D6.4 on the application of machine learning techniques to the inference of plant architecture from image or point cloud data. Here is a brief summary of the work done.

ROMI project requires to investigate advanced analysis and modelling techniques for inferring plant structures from various type of input data. A main issue that arises in this context is to predict a feature of the plant (species, health status, etc) from the observation of its branching architecture.

Here, our objective is to achieve this task by machine learning techniques. In general, while most of classification algorithms are only adapted to data living in Euclidean spaces (like scalars, vectors), trees do not admit a Euclidean representation. Consequently, we need to develop methods that are specific to tree data. Some classification algorithms, e.g., SVM, consists in mapping the input tree data into a given representation space via a so-called kernel function. Kernel functions adapted to trees, such as the subtree kernel, have been proposed by Vishwanathan and Smola in NIPS 2003. While the literature has

never been focused on the weight function involved in the subtree kernel, we have shown that this function is crucial in prediction applications. We thus have proposed a new algorithm for computing the subtree kernel. It has been designed to allow learning the weight function directly from the data. On some difficult datasets, we have improved the prediction error from 50% to 3%.

Finally, once the topologies of plants has been extracted from pictures taken by the scanning device, we will be able to apply our classification algorithm to these data.

4 Bibliography

Bensmihen, S., Hanna, A. I., Langlade, N. B., Micol, J. E. L., Bangham, A., & Coen, E. S. (2008). Mutational spaces for leaf shape and size. *Hfsp J*, 2(2), 110–120. <http://doi.org/10.2976/1.2836738>

Yann Guédon, Yassin Refahi, Fabrice Besnard, Etienne Farcot, Christophe Godin, and Teva Vernoux. Pattern identification and characterization reveal permutations of organs as a key genetically controlled property of post-meristematic phyllotaxis. *Journal of theoretical biology*, 338:94–110, 2013.

Lars Mündermann, Yvette Erasmus, Brendan Lane, Enrico Coen, and Przemyslaw Prusinkiewicz. Quantitative modeling of arabidopsis development. *Plant Physiology*, 139(2):960–968, 2005. ISSN 0032-0889. doi: 10.1104/pp.105.060483.

Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun. Open3D: A modern library for 3D data processing. arXiv:1801.09847, 2018.

Weight, C., Parnham, D., & Waites, R. (2008). TECHNICAL ADVANCE: LeafAnalyser: a computational method for rapid and large-scale analyses of leaf shape variation. *The Plant Journal*, 53(3), 578–586. <http://doi.org/10.1111/j.1365-313X.2007.03330.x>

S. Vishwanathan and A. J. Smola, “Fast Kernels for String and Tree Matching,” in NIPS, 2002, pp. 569–576.