OneRosette to Predict Them All: Single Plant Prompting on a Visual Foundation Model to Segment Symptomatic Arabidopsis Thaliana Time Series

Felicià Maviane Macia
Laboratoire des Interactions Plantes Micro-organismes Environnement (LIPME)
Sabine Wiedemann-Merdinoglu
University of Strasbourg
David Rousseau
david.rousseau@univ-angers.fr

Research Institute of Horticulture and Seeds
Nemo Peeters
Laboratoire des Interactions Plantes Micro-organismes Environnement (LIPME)

Method Article

Keywords: Deep learning, Visual foundation model, Arabidopsis thaliana, Pseudomonas syringae pv. tomato DC3000, Ralstonia pseudosolanacearum

Posted Date: May 7th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-6505262/v1

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Plant Methods on October 9th, 2025. See the published version at https://doi.org/10.1186/s13007-025-01432-2.

OneRosette to Predict Them All: Single Plant Prompting on a Visual Foundation Model to Segment Symptomatic *Arabidopsis Thaliana* Time Series

Felicià Maviane Maciá[1,2,3], Sabine Wiedemann-Merdinoglu[3], David Rousseau[2*†], Nemo Peeters[1*†]

[1] Laboratoire des Interactions Plantes Micro-organismes Environnement (LIPME), INRAE, CNRS, Université de Toulouse, Castanet-Tolosan, France .
[2] Institut de Recherche en Horticulture et Semences (IRHS), LARIS, INRAE, Université d'Angers, Angers, France .
[3] Santé de la Vigne et Qualité du Vin (SVQV), INRAE,Université de Strasbourg, Colmar, France .

*Corresponding author(s). E-mail(s): david.rousseau@univ-angers.fr; nemo.peeters@inrae.fr;
Contributing authors: felicia.maviane-macia@inrae.fr; sabine.wiedemann@inrae.fr;
†These authors contributed equally to this work.

Abstract

Background *Arabidopsis thaliana* is the leading model plant used to study plant-pathogen interactions. High-throughput phenotyping allows for the simultaneous study of many plants with high-frequency image acquisition. Nevertheless, the segmentation of symptomatic plants on natural soil remains challenging, requiring the annotation of hundreds of images and the subsequent training of specialized models for each pathosystem considered. This paper presents a novel approach to segmenting *A. thaliana* plants' time series using a single annotated image.

Results Images of *A. thaliana* plants infected with *Pseudomonas syringae* pathovar *tomato* strain DC3000 were annotated with precise segmentation masks. We compared various mask segmentation methods; our one-shot learning approach obtained a Dice score of 0.977 on our test dataset. Variables extracted from the segmented images allowed statistical discrimination between infected and control plants. We used our one-shot learning approach without further fine-tuning on a new pathosystem; *A. thaliana* infected with *Ralstonia pseudosolanacearum*, strain GMI1000. We obtained a Dice score of 0.966 in the second test dataset. We also obtained a Pearson correlation coefficient of -0.928 between the annotated quantitative disease index and the variable generated with our method.

Conclusion This work provides a pipeline to segment symptomatic *A. thaliana* plants by leveraging a visual foundation model. The method has been used successfully on two different pathogens, is fast to train, and does not need a large dedicated graphical processing unit. Our method has characterized

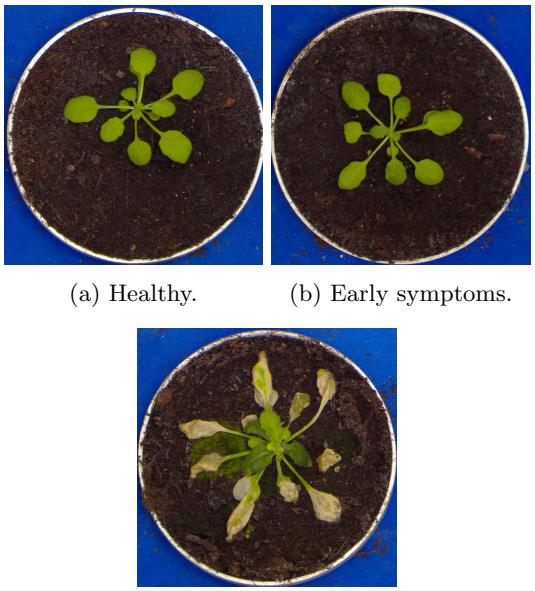
plant-pathogen interactions of two pathosystems without fine-tuning for the second pathosystem. Its ease of use and low computing requirements should make adapting our approach to other high-throughput phenotyping platforms easy.

Keywords: Deep learning, Visual foundation model, *Arabidopsis thaliana*, *Pseudomonas syringae* pv. *tomato* DC3000, *Ralstonia pseudosolanacearum*

1 Background

Arabidopsis thaliana is one of the most-studied plants in contemporary biology [1] despite not having an agronomic interest. Indeed, its small footprint, ease of growth, availability of genomic and other resources, and susceptibility to multiple pathogens make it an ideal candidate for studying plant growth [2], the effect of biotic [3, 4], or abiotic stresses [5]. One such biotic stress is *Pseudomonas syringae*, a bacterial pathogen that can infect various plants, including many crops [6]. Among its pathovars, *P. syringae* pathovar *tomato* strain DC3000 (*Pst* hereafter) has been found to infect various *A. thaliana* ecotypes [7]. This discovery has made the *A. thaliana*-*Pst* pathosystem the most used model to study plant-pathogen interactions, both plant resistance mechanisms and pathogen fitness properties. The pathogen's leaf symptoms consist of stains with colors ranging from yellow to brown.

In recent years, high-throughput phenotyping, in combination with deep learning techniques, as an alternative to manual phenotyping, has improved our ability to conduct studies on plant-pathogen interactions. This approach is faster when using many plants, removes human bias, thus improving consistency, and yields far more parameters. However, deep learning methods require many annotated images. Furthermore, each new pathosystem requires new annotations and training. Another essential drawback is that segmentation performance decreases when the foreground and background have similar visual properties, which is the case when plants are infected with leaf pathogens, as seen in Fig. 1.



(a) Healthy. (b) Early symptoms.

(c) Late stage.

Fig. 1: Healthy *A. thaliana* sample plant (a) and two samples of *A. thaliana* with *Pst* symptoms: one for early stage (a) and one for late stage (c).

This study proposes a novel approach to segmenting time series of symptomatic *A. thaliana* plants on soil using a single annotated image. We present OneRosette, a pipeline that leverages a visual foundation model using a one-shot

learning approach. We obtained good segmentation scores even when plants presented advanced symptoms. We validated the pipeline's usability to describe plant-pathogen interactions by performing two experiments using color, morphology, and texture features extracted from the segmented images. In the first experiment, we assessed how early we could detect *Pst* symptoms in infected *A. thaliana*. In the second experiment, we used the same pipeline with no further fine-tuning to track the symptoms' evolution of *A. thaliana* plants infected with a second bacterial pathogen, *Ralstonia pseudosolanacearum* (*Rps* hereafter).

2 Related work

To our knowledge, there are no studies on the *A. thaliana-Pst* interaction with whole plants grown on soil using image processing phenotyping. One of the reasons for this lack of studies is the added difficulty of segmenting symptomatic *A. thaliana* rosettes on soil.

When automatically phenotyping *A. thaliana*-disease interactions, a common approach was to study excised leaves on uniform backgrounds. In [8], the authors study the interaction between detached leaves on a blue background infected with *Botrytis cinerea* after their excision for up to 96 hours after infection. This method allows for easy segmentation, but we can't exclude that excised leaves behave differently from their plant-attached counterparts. Furthermore, the study duration is limited to the time that a detached leaf can survive. In [9], the authors used excised leaves to study a bacterial pathogen, *Xanthomonas campestris pv. campestris*. These two articles show a common problem when using machine learning techniques: the need for separate tools for each pathogen studied.

AraDQ [10], a digital phenotyping software, allows the study of *A. thaliana* rosettes grown on agar Petri dishes and infected with *Pst*. In this work, segmentation used a YOLO v4 [11] model and needed 1996 annotated images for training. The disease was evaluated using 10 image parameters, 4 for colorimetry and 6 for morphology. The pipeline was then validated with two case studies. This is an improvement, as it allows whole-plant phenotyping. However, there is still a need for a large amount of annotated images, and importantly, the requirement of in vitro settings prevents any actual soil plant conditions.

There are no studies or datasets related to symptomatic *A. thaliana* plants; however, in the Computer Vision Problems in Plant Phenotyping (CVPPP) challenge dataset [12–14] mask-annotated images of *A. thaliana* plants and *Nicotiana benthamiana* plantlets are present. These datasets are the *de facto* benchmark for healthy *A. thaliana* plant segmentation. Improvements in this data set since 2014 have led to the current

state of the art described in [15] that uses a U-Net architecture [16] with a MobileNetV2 encoder [17].

In contrast to existing methods, our pipeline can segment healthy and symptomatic plants growing on soil using only a single annotated image. This study used two bacterial pathogens that cause various symptoms: wilting, chlorosis, and leaf spots of diverse colors, textures, and shapes affecting the healthy *A. thaliana* leaves.

3 Methods

3.1 Image Acquisition

PHENOPSIS [18] is a fully automated growth chamber that regulates temperature, lighting, watering, and hygrometry. The chamber can accommodate 504 plants in 14 trays containing 36 pots each. An RGB line scanner takes top-view photos of the plants multiple times daily; the images have a 3337x3337 pixel resolution.

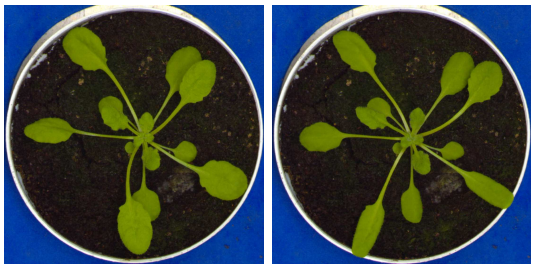
Let us define the following abbreviations: D{XX} is the XX day after sowing. D{XX}-{YY}h refers to the timestamp of the image acquisition cycle that started the XX day after sowing at YY hours.

3.2 Plant Material

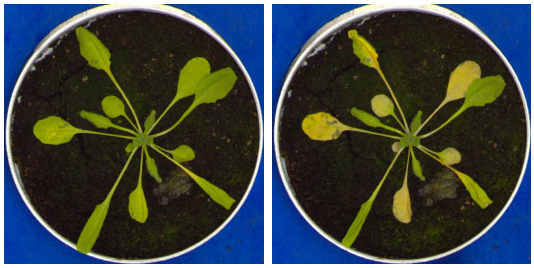
All *A. thaliana* were of ecotype Columbia (Col-0). All experiments consisted of 144 plants. Seeds were sown at D00 in individual pots 72mm in diameter and grown for the entire length of the experiments inside the PHENOPSIS chamber. The pots were filled with long Jiffy "forestry" pellets soaked in tap water overnight. The Jiffy inside the phenopsis pots were topped up with PAM 2 Proveen's potting soil. These pots were then gently sprayed on the top with water and seeded. The target weight watering was set at 172g per individual pot. Each pot was weighed and watered daily at midnight.

We performed two separate experiments with two bacterial pathogens. In the first experiment, we used *Pst* DC3000, a bacterial pathogen capable of infecting *A. thaliana* plants. Observed leaf symptoms included water-soaked patches (Fig. 1b) and later necrosis with chlorosis (Fig. 1c) [19]. We used 144 plants grown on an 8/16h day/night cycle, with day temperatures set at 22°C and night temperatures at 21°C. Image acquisition for all plants started on D18 at 12 hours and ended on D32 at 15 hours. Image acquisition cycles were performed twice daily at 12 and 15 hours. We obtained 4032 images. After D21-12h, when the plants were at a development stage suitable for infection, the plants were divided into two equal-sized groups, infected and control, using stratified randomization on the tray. The infected group plants were inoculated with a spray containing a solution at 10^{-7} cfu/ml in 10mM MgCL2 with $0,04\%$ Silwet. Each plant was inoculated with three sprays, allowing a complete leaf area

coverage. After the infection, all plants were covered with small (60 mm diameter) Petri dishes for 22 hours to facilitate infection. While plants were covered, acquisition cycles were paused.



(a) *Rps* DI 0. (b) *Rps* DI 1.



(c) *Rps* DI 2. (d) *Rps* DI 3.

(e) *Rps* DI 4.

Fig. 2: *Rps* sample images for all disease index values.

In the second experiment, we used *Rps*, the causal agent of bacterial wilt, a second bacterial pathogen capable of infecting *A. thaliana* plants. The first infection symptom was leaf wilting (Fig. 2b), then, yellow stains progressively appeared (Figures 2c and 2d, and the whole plant finally died and turned brown (Fig. 2e). In this experiment, we used 144 plants grown on an 8/16h day/night cycle, with day temperatures set at 23°C and night temperatures at 22°C. One day before infection (D26), we raised the temperatures to 28°C during the day and 27°C at night, the optimal pathogen infection temperature. Like the previous experiment, image acquisition started at D18 at 13 hours and ended at D40 at 17 hours. Image acquisition cycles were performed twice daily at 13 and 17 hours, producing 6336 images. After D27-12h, when the plants were in a development state suitable for infection, they were divided into two equal-sized groups, infected and control, using stratified randomization on the tray. The infected group plants were inoculated, by watering, with 20ml per plant of an *Rps* bacterial solution at 10^8 cfu/ml in tap water. We added 20 ml of water to the control plants to equilibrate for the added water weight. We did not need to cover the plants or pause the acquisition cycles for this experiment.

3.3 Proposed Method

As seen in Fig. 3, the proposed pipeline has five steps. The source images were cropped (i) to the main pot. The first cropped image of the time series was then thresholded (ii) to obtain a three-class image, which was used to select both plant and background prompts (iii). The cropped image

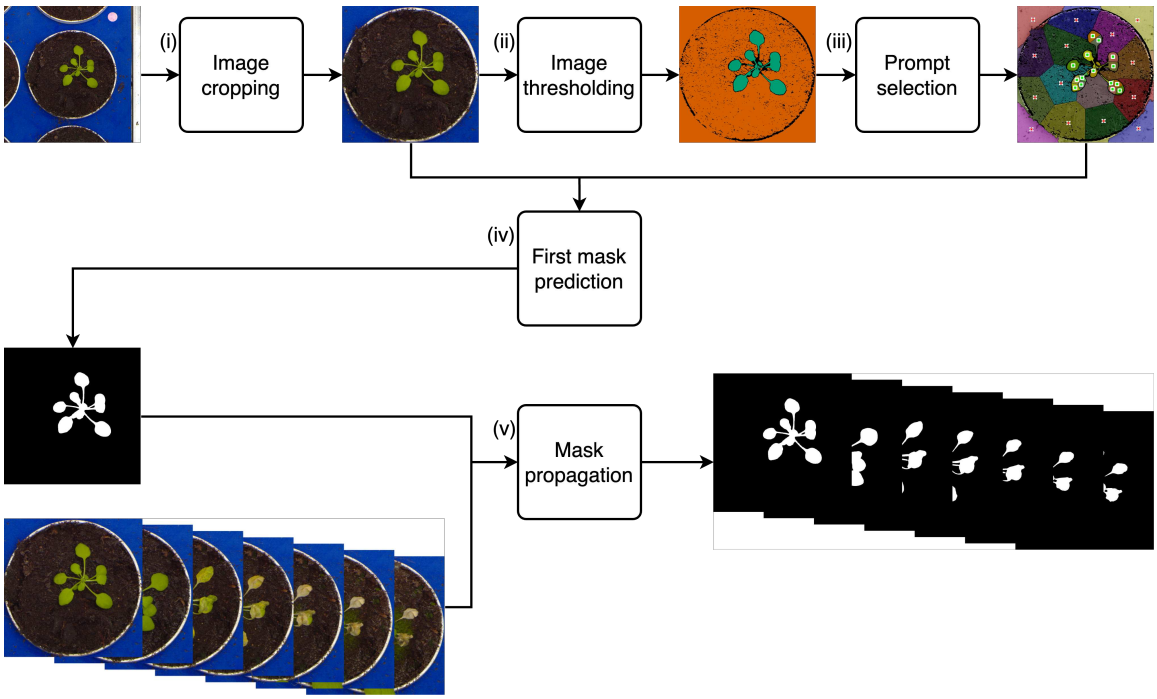


Fig. 3: Time series segmentation pipeline.

and the prompts were then used to predict the first image's mask (iv), which was propagated (v) to the rest of the time series' cropped images. The method has three hyperparameters: a percentile q for thresholding and the number of plant and background prompts. Hereafter, a detailed description of the method's steps.

Image Cropping

The images generated by the high-throughput platform were 3337x3337 pixels in RGB format. They contained one pot at the center, of a diameter of 2200 pixels, and parts of neighboring pots and plants on the borders. The position of the central pot in the image could vary by up to 128 pixels on the x-axis and 176 pixels on the y-axis. To avoid issues when segmenting the time series, we cropped the source images to the central pot using a Hough circle detector [20] as implemented in the OpenCV library [21]. We used a Canny filter [22] with a Sobel kernel size of three on the RGB blue channel as input for the Hough detector to take advantage of the contrast between the blue background, the white pot, and the brown soil. We used lower and upper hysteresis thresholds of 0 and 255 while looking only for circles with diameters between 2100 and 2200 pixels to improve the chances of finding the pot. The detected circles had an average diameter of 2124.96 pixels; to ensure that all resulting images had the same size, we resized all circles to 2200 pixels before cropping to the bounding square.

Image Thresholding

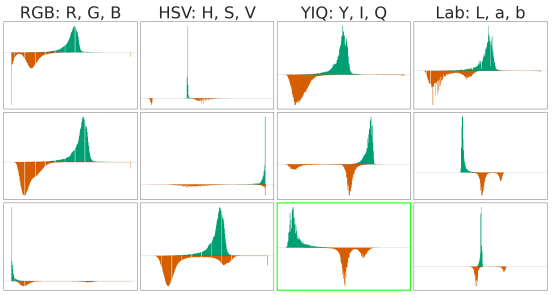


Fig. 4: Separated accumulated histograms for all channels of all color spaces. Green histogram for plant pixels and orange histograms for background pixels. The best histogram is highlighted in green.

The next step was to classify the pixels of the first image in the time series into one of three classes: plant, background, and unknown. To do this, we used all available images in the dataset with their segmentation masks to calculate separate accumulated histograms for the plant and background pixels for each channel of four color spaces: RGB, HSV, YIQ, and Lab [23]. Let $P_{cs,cn} = (p_0, ..., p_{255})$ and $B_{cs,cn} = (b_0, ..., b_{255})$ respectively be the plant and background pixels histograms for the channel cn of the cs color space. We calculated the Hellinger distance

$$H(P_{cs,cn}, B_{cs,cn}) = \frac{1}{\sqrt{2}} \sqrt{\sum_{i=0}^{255} (\sqrt{p_i} - \sqrt{b_i})^2}, \quad (1)$$

for all channels of all color spaces, where higher $H(P_{cs,cn}, B_{cs,cn})$ values indicate less overlap between the histograms. Figure 4 shows histograms for all channels of all color spaces with a green highlight for the channel with the highest Hellinger distance.

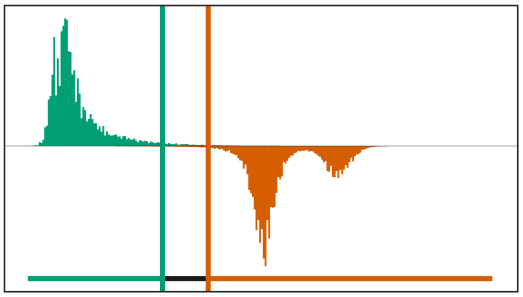
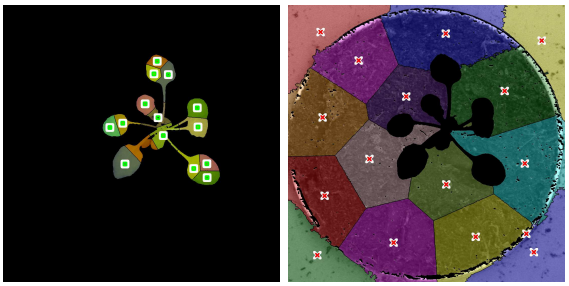


Fig. 5: Selected histogram with plant (green) and background (orange) thresholds. The bottom horizontal bar shows the ranges for each class: plant in green, unknown in black, and background in orange.

Figure 5 shows an example of a histogram selected with Eq. 1. In this example, pixels with low values belong primarily to the plant, and pixels with higher values belong mainly to the background. A small portion of the pixels in the middle had an equal probability of belonging to the plant or the background. We positioned two thresholds using the selected histogram to classify the cropped image pixels into the three classes automatically. The first threshold was positioned at the percentile b of the background histogram, with $0 < b < 0.5$, green vertical bar in Fig. 5, pixels with values under the threshold were assigned to the plant class. The second threshold was positioned at the percentile p of the plant histogram, with $0.5 < p < 1$, orange vertical bar in Fig. 5, pixels with values over the threshold were assigned to the background class. The pixels between the two

thresholds had a similar probability of belonging to the plant and the background and were assigned to the unknown class. To simplify the method, we derived both percentiles from a single value q, such as $b = q$ and $p = 1 - q$. This percentile q was the first method's hyperparameter we needed to fine-tune.

Prompt Selection



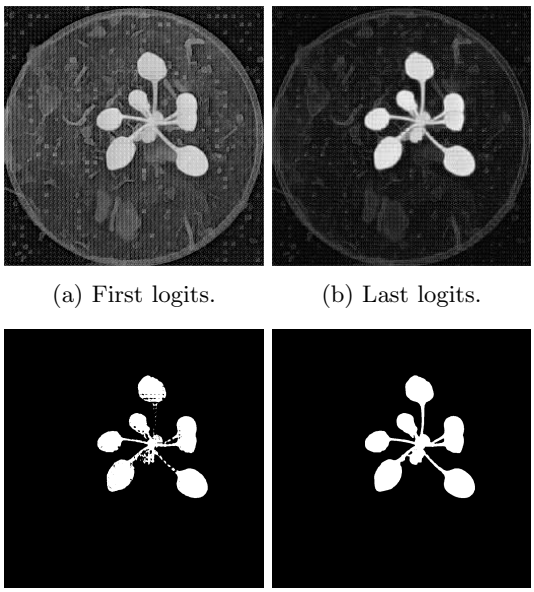
(a) Plant prompts. (b) Background prompts.

Fig. 6: Plant (a) and background (b) prompts were extracted from the plant and background pixels images using the SLIC method.

Once the image pixels were classified into the three classes, we selected prompts from pixels belonging to the plant and background classes. The SAM v2 [24] Visual Foundation Model uses images and prompts to predict segmentation masks. The prompts can be bounding boxes, masks, or points. We chose the points because they allowed for creating plant and background prompts. The number of plant and background prompts were the two other methods' hyperparameters that needed fine-tuning. We treated the classified plant and background pixels as separate images. We applied the SLIC method [25] with the number of segments equal to the corresponding prompts count. We chose a SLIC compactness of 50 to prioritize shape over color, as foreground and background colors are homogeneous in the first image of a time series. From each segment generated with SLIC, we extracted the point of inaccessibility [26] as implemented in polylabel [27], which, contrary to the centroid, is guaranteed to always be within the shape even if they are concave. Figure 6 shows an example of SLIC segments and the points of inaccessibility used as prompts for plant and background pixels. Points extracted from the plant pixels were used as plant prompts, and the ones extracted from background pixels were used as background prompts.

First Mask Prediction

Once we had the cropped image and prompts for the first image of the time series, we used them as input for the SAM v2 model. The model outputs three objects: the predicted mask, the mask's confidence score, and the mask's logits, a low-resolution unthresholded version of the mask. Once the first logits are predicted, SAM v2 can be used iteratively by supplying the previous iteration's mask's logits as input for the next iteration. We used the mask's score to monitor the mask's evolution and stopped iterating when the score between two iterations improved by less than



(a) First logits. (b) Last logits.

(c) First mask. (d) Last mask.

Fig. 7: Comparison between first and final logits (a, b) and mask (c, d).

0.001. Figure 7 shows an example of the logits and mask for the first and last iterations.

Mask Propagation

Once we had cropped images for all photos in the time series and the first mask, we leveraged SAM v2's video segmentation capabilities to propagate the first predicted mask throughout the whole time series and obtain masks for each image.

3.4 Datasets

In this study, we performed two types of experiments. To validate the OneRosette accuracy at a pixel level, we conducted a segmentation experiment on *A. thaliana* plants infected with *Pst* and *Rps*. Then, to assess OneRosette's performance at a time series level, we performed two types of disease tracking experiments: in the first one, we tried to detect disease onset as early as possible, and in the second, we aimed at tracking the disease's evolution. This needed multiple datasets of different types, which prompted us to create a naming convention to distinguish them easily. The names of datasets can consist of up to four parts: (i) **Pathogen**, either "*Pst*" or "*Rps*", (ii) **Type** either "segmentation" when images are annotated with segmentation masks, "quantitative" when annotated data is based on a disease index, or "qualitative" when the symptom's presence is annotated as a binary value, (iii) **Subset** either "training", "validation", or "test", and (iv) **Composition** (only for segmentation datasets), either "full", when the dataset contains control and infected plants, "healthy", when only control plants are present, and "one-shot" when the dataset contains a single image from the first acquisition cycle.

Figure 8 shows an overview of all the datasets from experiments with this pathogen. We started by creating the *Pst* Qualitative dataset, where an expert annotated whether there were or were not any visible symptoms (i.e., the smallest visible symptom on at least one leaf was considered); the control plants' images were automatically annotated as not having any symptoms. Figure 9 shows the evolution of the number of symptomatic plants over time. All infected plants were symptomatic at D24-12h, three days after infection. This first

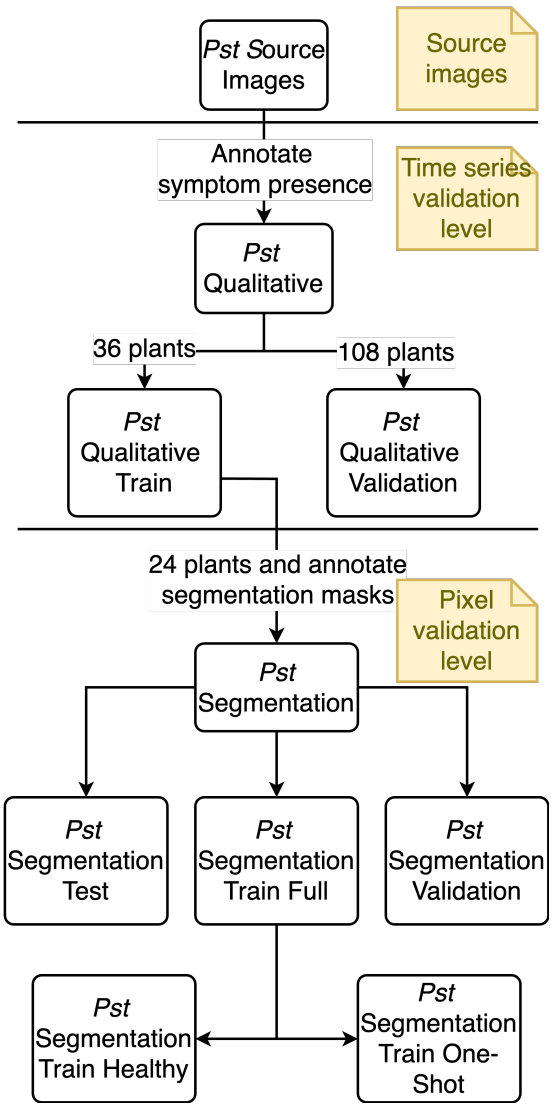


Fig. 8: Overview of the *Pst* datasets.

dataset was split into two subsets: Train, which had 36 plants, and Validation, which had 108 plants. Next, we randomly selected the images of six control and 18 infected plants from the *Pst* Qualitative Train dataset. We annotated them with segmentation masks using Make Sense [28], a mask segmentation annotation software, to create the *Pst* Segmentation dataset and its descendants.

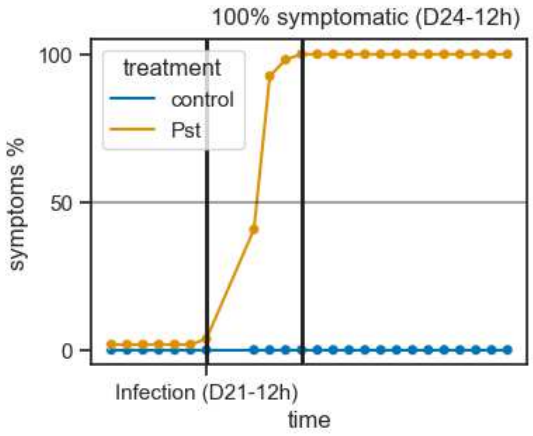


Fig. 9: Evolution of the percentage of *Pst* infected plants with symptoms. After infection (after D21-12h).

The resulting dataset contained 672 annotated images and was split into three subsets: Train, containing 476 photos from 17 plants, Validation, containing 84 images from three plants, and Test, containing 114 images from four plants. To assess the impact of the size and the composition of the training dataset on the different approaches' performance, we created three datasets with different compositions:

- *Pst* Segmentation Train Full contains the entire training dataset, 476 images from 17 plants, 4 control, and 13 infected plants.
- *Pst* Segmentation Train Healthy includes 112 images from four control plants from the training dataset.
- *Pst* Segmentation Train One-shot contains the first image in the time series of each plant in the *Pst* Segmentation Train Full dataset.

All images in this dataset were taken before infection and are therefore asymptomatic.

Table 1: Image distribution for the *Pst* segmentation datasets.

Countss	Train	Validation	Test
	Images(plants)		
Full Dataset	476(17)	84(3)	112(4)
Healthy Dataset	112(4)	-	-
One-Shot Dataset	17(17)	-	-

Table 1 shows an overview of all the *Pst* segmentation datasets.

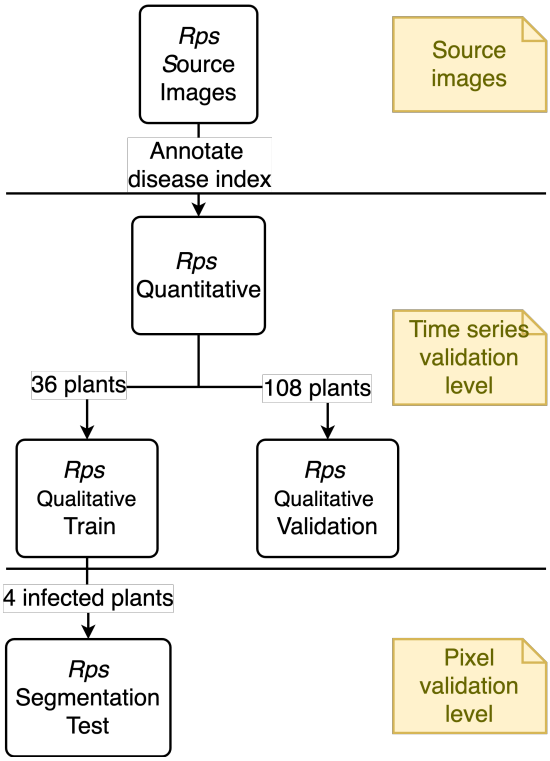


Fig. 10: Overview of the *Rps* datasets.

We then created the datasets related to the *Rps* pathogen. Figure 10 shows an overview of all the datasets created with images infected with this pathogen. For the *Rps* Quantitative dataset, an expert annotated all the infected plants' images with a disease index score [29] with values ranging from 0 to 4 with the following code:

0. No symptoms, Fig. 2a.
1. The first day displaying at least one wilted leaf can be more than one, Fig. 2b.
2. Half of the rosette leaves are wilted, and yellow stains may appear, but the rosette can also be green overall, Fig. 2c.
3. Almost all leaves are wilted, and yellow stains appear. Young leaves remain green, Fig. 2d.
4. All leaves, including the young ones, are wilted. Yellow, brown, and black stains cover most of the plant, Fig. 2e.

All images from the control plants were automatically assigned a disease index of 0.

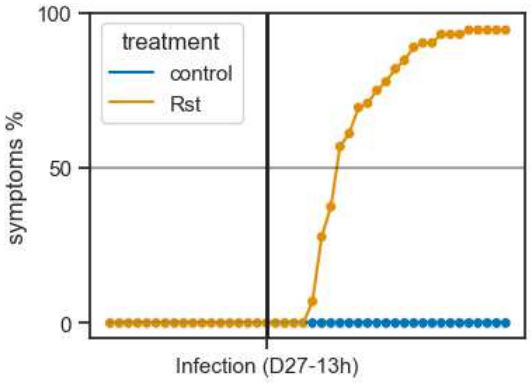


Fig. 11: Evolution the percentage of *Rps* infected plants with symptoms.

Figure 11 shows the evolution of the percentage of symptomatic plants over the experiment's duration. The vertical black bar marks the infection timestamp. Compared to the *Pst* experiment,

symptoms took longer to appear post-infection, 27 hours for *Pst* symptoms and 52 hours for *Rps* symptoms. At the end of the experiment, four infected plants remained asymptomatic. The dataset was split into two subsets: Train and Validation. Table 2 shows the number of images

Table 2: Count of annotated images for each disease index score.

disease index	0	1	2	3	4
count	5133	138	251	289	525

for each disease index in the annotated dataset. We selected the images from four plants of the *Rps* Quantitative Train dataset to create the final dataset and annotated them with segmentation masks to create the *Rps* Segmentation Test dataset.

3.5 Pixel Level Validation

We used two metrics to assess the different models' performances when segmenting images:

$$Dice(U, V) = \frac{2 * |U \cap V|}{|U| + |V|} \tag{2}$$

and Intersection over Union (IoU)

$$IoU(U, V) = \frac{|U \cap V|}{|U \cup V|} . \tag{3}$$

Both metrics evaluate segmentation, but IoU penalizes more errors, thus being a good indicator of the worst-case scenario.

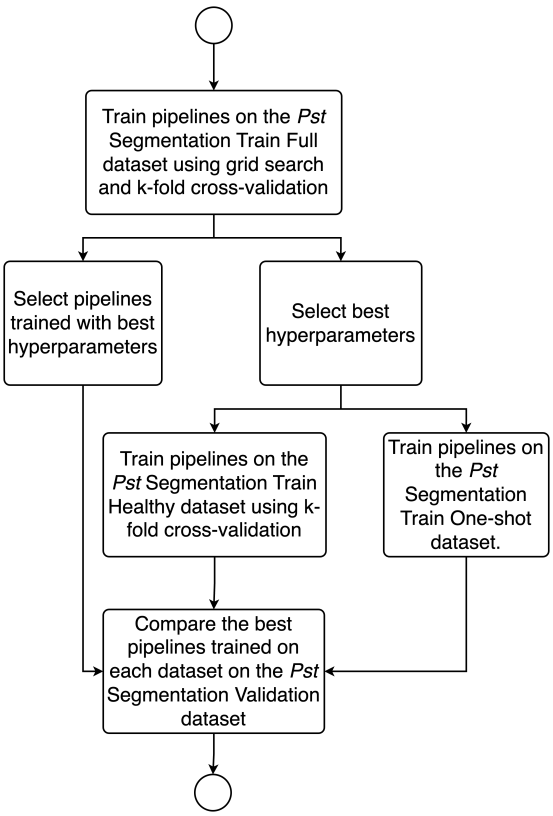


Fig. 12: OneRosette time series segmentation hyperparameter fine-tuning process.

We fine-tuned the three hyperparameters: q-percentile, plant prompts, and background prompts using 5-fold cross-validation coupled with a grid-search on the *Pst* Segmentation Train Full dataset. Folds were created to guarantee that all the images from a plant would always be in the same fold. We tested the following values for each hyperparameter:

- Histogram percentile: 0.0005, 0.001, 0.005, 0.01.
- Plant prompt count: 32, 64, 128, 256.

- Background prompt count was defined relative to plant prompt count with values none, half, same, or double.

We used the best hyperparameters to train pipelines on the two other *Pst* Train datasets. We used 4-fold cross-validation for the healthy dataset, one less fold than for the Full dataset because the Healthy dataset contained only four plants. We also trained pipelines on each image in the *Pst* Segmentation Train One-shot dataset. We validated them on the hold-out *Pst* Segmentation Validation dataset as a validation dataset.

In the absence of previous studies on our dataset, we used a neural network composed of a U-Net [16] architecture with a MobileNetV2 encoder [17] pre-trained on ImageNet [4] as a baseline comparison to our approach. The models were implemented using Segmentation Models Pytorch [30], and training was performed using Pytorch Ignite [31]. We performed image augmentation using the Albumentations [32] library. All images were resized to 1024x1024 pixels using linear interpolation. Then, a series of augmentations was randomly applied each time an image was sent to the network. The selected augmentations were rotations, vertical and horizontal symmetries, brightness and contrast modifications with limits of 0.2 and 0.2, respectively, random gamma with upper and lower bounds of 0.8 and 1.2, and contrast limited adaptive histogram equalization. For the training, we selected a batch size of 27 to fill as much as possible the GPU's RAM, an NVIDIA A100 80G, and a learning rate of 0.001. The models were trained for a maximum of 500 epochs with an early stopping monitoring of the Dice loss and a waiting time of 10 epochs. We trained the model 10 times with each training dataset: Full, Healthy, and One-shot. In all cases, we used the hold-out *Pst* Segmentation Validation dataset as a validation dataset.

Once all pipelines and models were trained, we compared the best OneRosette pipelines and U-Net neural networks trained on each of the three *Pst* Segmentation Train datasets' performances on the hold-out *Pst* Test dataset. We also compared the time needed to train each approach on each dataset.

Finally, to assess the method's performance when changing the pathosystem, we compared the scores of the best OneRosette pipelines and U-Net neural networks trained on each of the three *Pst* Segmentation Train datasets on the hold-out *Rps* Segmentation Test dataset.

3.6 Time Series Level Validation

To validate the use of our pipeline trained on a single image on biological studies, we performed experiments using images of *A. thaliana* plants infected with two pathogens *Pst* and *Rps*.

We segmented all available images with our one-shot pipeline. Then, we extracted mean and standard deviation values for each channel of

the four color spaces. We also extracted morphological features: area, bounding box area, extent, convex hull, solidity, perimeter, and minimum enclosing circle radius. Finally, we used gray-level co-occurrence matrices [33] (GLCM) to extract contrast, dissimilarity, homogeneity, angular second moment (ASM), energy, and correlation from the segmented images. In the images with symptomatic plants, the transition between symptomatic and asymptomatic areas was between 2 and 10 pixels wide. To build the GLCM, we selected a distance of 10 pixels, which captured all transitions. We selected angles at 0°, 45°, 90°, and 135° with a symmetric matrix and averaged results over angles because, in our case, symptoms are rotationally invariant. This process created 24 color, 7 morphological, and 6 texture variables for a total of 37 variables.

We then used the *Pst* Qualitative and the *Rps* Quantitative datasets to find the variables that best described the plant-pathogen interaction for each pathosystem. To prevent the different magnitudes of each variable from influencing the selection process, we started by applying a min-max feature scaling

$$v' = \frac{v - v_{min}}{v_{max} - v_{min}} \tag{4}$$

to all variables.

Let

$$\overline{v'_T}(P_x) = \frac{\sum_{p_x}^{P_x} \sum_t^T v'_t(p_x)}{|T| * |P_x|} \tag{5}$$

be the average value of the normalized variable $v' \in V'$, the variables extracted from the segmented images, over T timestamps of all the plants $p_x \in P_x$ where x the treatment is either control (c) or infected (i). The selected variable should satisfy two conditions: It should minimize

$$\Delta_{v'_{t=0}} = |\overline{v'_{t=0}}(P_c) - \overline{v'_{t=0}}(P_i)| \tag{6}$$

the average distance between the values for control and infected plants for the variable v at timestamp 0 between the control and infected plants, and it should maximize

$$\Delta_{v'_{t>0}} = |\overline{v'_{t>0}}(P_c) - \overline{v'_{t>0}}(P_i)| \tag{7}$$

the average distance between the values for control and infected plants for the variable v for timestamps after infection between the control and infected plants. Eq. 6 filtered variables that may have different values before infection. At the same time, Eq. 7 selected the variables with the largest distance between control and infected plants during the symptom onset phase. As a consequence, we selected v_s, the selected variable, the variable that satisfied

$$v_s = \underset{v' \in V'}{\operatorname{argmmax}}\, f(v') \tag{8}$$

with

$$f(v') = \Delta_{v'_{t>0}} - \Delta_{v'_{t=0}} \quad (9)$$

as the variable that best describes the difference between the infected and control plants. Some variables may be highly correlated, which may cause Eq. 9 to yield minor differences between the top variables. To avoid this issue, we used bootstrap sampling a thousand times on the dataset, calculated $f(v')$ for all variables, and selected the variable that came first on more occasions.

In the first experiment, we used the *Pst* images to assess how early we could detect a significant difference between the control and infected plants. For that, we used the images acquired between D21-12h, the infection timestamp, and D24-12h, the moment when all infected plants were symptomatic in the *Pst* Qualitative Train dataset to select the variable that satisfied Eq. 8. We calculated ANOVAs at each timestamp with data extracted from the *Pst* Qualitative Validation dataset. Our H_0 hypothesis was that the means of the control and infected plants for v_s were the same. We added the tray on which the plant was positioned as a second factor to the ANOVAs and checked whether there was an interaction between the treatment and the plant's position.

Similarly to the first experiment, we used the *Rst* Quantitative Train dataset images taken between D27-13h and D40-13h to select the variable that satisfied Eq. 8. We then performed the same ANOVAs with data from the *Rst* Quantitative Validation dataset. Finally, we assessed the possibility of using the selected variable as a disease evolution indicator by calculating the Pearson correlation coefficient between the annotated disease index and the selected variable.

4 Results

4.1 Pixel Level Validation

Tables 3 and 4 show Dice and IoU score means with standard deviation when using cross-validation and grid-search for hyperparameter selection. The best Dice score is 0.9803 with a standard deviation of 0.0008. The best IoU score is 0.9615 with a standard deviation of 0.0015. We obtained the best scores in both metrics using a q-percentile of 0.005, 128 plant prompts, and 256 background prompts. Scores are similar except when using no background prompts and 128 or more plant prompts, in which case, metrics drop under 0.4 for Dice and IoU. Table 5 shows results when comparing pipelines trained on each *Pst* Segmentation Train dataset on the hold-out *Pst* Segmentation Validation dataset. The pipelines have identical IoU scores and a minimal difference of 0.001 in favor of the pipeline trained on the one-shot dataset for the Dice score.

When analyzing the results of training the U-Net model with all three datasets (Tab. 6), we

Table 3: Mean Dice with standard deviation for the cross-validation when fine-tuning hyperparameters with grid-search. Best hyperparameters and their score in bold font.

	Background prompts	Dice			
		None	Half	Same	**Double**
Percentile	Plant prompts				
0.0005	32	0.93±0.08	0.95±0.06	0.90±0.16	0.91±0.09
	64	0.92±0.06	0.92±0.07	0.96±0.04	0.93±0.09
	128	0.33±0.11	0.975±0.003	0.9785±0.0023	0.9786±0.0018
	256	0.25±0.05	0.9727±0.0016	0.9784±0.0017	0.979±0.003
0.001	32	0.9740±0.0017	0.9781±0.0010	0.979±0.003	0.9796±0.0014
	64	0.9702±0.0018	0.9776±0.0025	0.9795±0.0011	0.9796±0.0021
	128	0.41±0.18	0.9777±0.0019	0.9797±0.0026	0.9797±0.0019
	256	0.25±0.08	0.976±0.003	0.9798±0.0014	0.9796±0.0012
0.005	32	0.9741±0.0018	0.9785±0.0019	0.9800±0.0025	0.9798±0.0013
	64	0.9694±0.0020	0.9780±0.0012	0.9800±0.0024	0.9799±0.0017
	128	0.32±0.10	0.9771±0.0019	0.9803±0.0013	**0.9803±0.0008**
	256	0.26±0.04	0.9753±0.0016	0.9799±0.0011	0.9802±0.0011
0.01	32	0.9739±0.0013	0.9780±0.0027	0.9795±0.0024	0.9796±0.0020
	64	0.9685±0.0016	0.9776±0.0020	0.9799±0.0015	0.9799±0.0018
	128	0.36±0.14	0.9767±0.0015	0.9801±0.0013	0.9802±0.0016
	256	0.25±0.06	0.9751±0.0022	0.9798±0.0014	0.9794±0.0018

observed that the model's performance was optimal when trained with the full training dataset, and scores decreased when removing images from symptomatic plants. The scores decreased even further when the one-shot dataset was used to train the model. As seen in Tab. 7, our approach is around 35 times faster to train than a U-Net when using the full dataset.

We used the validation scores to select the best U-Net models and the best OneRosette pipelines and compared the metrics scores on the hold-out *Rps* Test Dataset. As shown in Tab. 8, the best results were obtained with the U-Net model trained on the Full Training Dataset. The OneRosette pipelines came behind with a difference in Dice score between 0.005 and 0.006, and the difference for IoU of 0.01. U-Net networks trained on the Healthy and One-shot datasets obtained the worst scores.

We used a symptomatic image from the last acquisition timestamp, Fig. 13a, to compare the predictions of our pipeline trained on the One-shot dataset to the U-Net models. The ground truth to which all predictions were compared is shown in Fig. 13b. Our pipeline, Fig. 13c, made the best prediction with a Dice score of 0.985; errors consisted of small over-segmentation near the plant's core, no moss was mistaken as part of the plant, and no symptomatic leaves were missed.

Table 4: Mean IoU with standard deviation for the cross-validation when fine-tuning hyperparameters with grid-search. Best hyperparameters and their score in bold font.

	Background prompts	IoU			
		None	Half	**Same**	**Double**
Percentile	Plant prompts				
0.0005	32	0.90±0.10	0.91±0.08	0.87±0.20	0.88±0.11
	64	0.88±0.09	0.89±0.09	0.92±0.05	0.90±0.11
	128	0.22±0.12	0.951±0.006	0.958±0.004	0.958±0.003
	256	0.16±0.04	0.947±0.003	0.958±0.003	0.958±0.006
0.001	32	0.949±0.003	0.9572±0.0020	0.959±0.006	0.9600±0.0027
	64	0.942±0.003	0.956±0.005	0.9600±0.0021	0.960±0.004
	128	0.32±0.18	0.957±0.004	0.960±0.005	0.960±0.004
	256	0.155±0.059	0.953±0.006	0.9604±0.0025	0.9602±0.0022
0.005	32	0.950±0.003	0.958±0.004	0.961±0.005	0.9605±0.0025
	64	0.941±0.003	0.9571±0.0024	0.961±0.005	0.961±0.003
	128	0.21±0.10	0.955±0.003	0.9614±0.0025	**0.9615±0.0015**
	256	0.16±0.04	0.952±0.003	0.9606±0.0022	0.9611±0.0021
0.01	32	0.9493±0.0024	0.957±0.005	0.960±0.005	0.960±0.004
	64	0.939±0.003	0.956±0.004	0.9607±0.0029	0.961±0.004
	128	0.26±0.14	0.9545±0.0029	0.9611±0.0024	0.961±0.003
	256	0.15±0.04	0.952±0.004	0.9605±0.0027	0.960±0.003

Table 5: Dice and IoU scores on the hold-out validation dataset when training OneRosette pipelines with the best hyperparameters.

Train dataset	Dice	IoU
Full	0.980±0.006	0.962±0.010
Healthy	0.980±0.006	0.962±0.010
One-shot	0.981±0.006	0.962±0.010

Table 6: Dice and IoU scores on the hold-out validation dataset when training U-Net neural networks on the three datasets.

Train dataset	Dice	IoU
Full	0.9850±0.0005	0.970±0.001
Healthy	0.937±0.010	0.891±0.014
One-shot	0.85±0.16	0.77±0.22

The U-Net model trained on the Full Dataset, Fig. 13d, obtained a Dice score of 0.967 with errors of over-segmentation of moss near the plant. The U-Net models trained without symptomatic plants

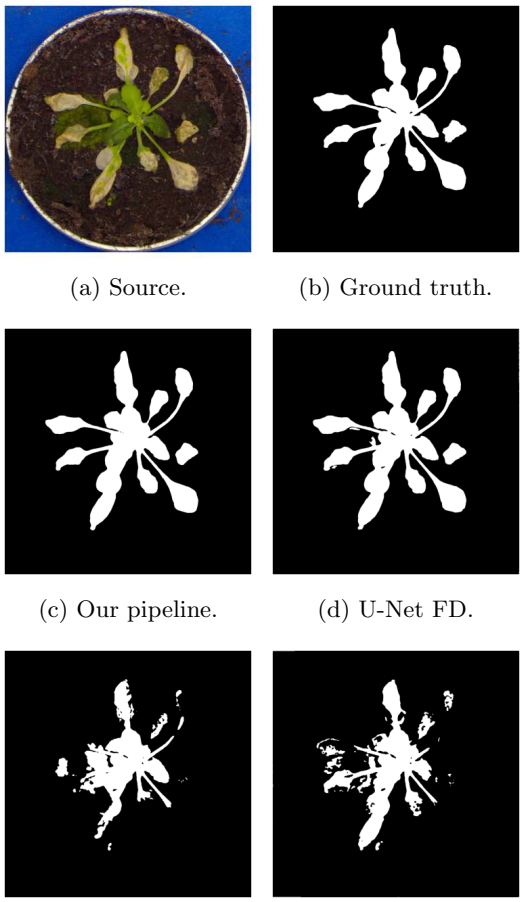
Table 7: Average time to train each approach on each dataset.

Dataset	U-Net	OneRosette
Full	1:46:13	3:10
Healthy	18:56	2:15
One-Shot	9:31	1:14

Table 8: Dice and IoU with standard deviation on the hold-out *Pst* segmentation test dataset for the best models of each approach trained on each dataset.

Method	Dataset	Dice	IoU
OneRosette	Full	0.976±0.010	0.954±0.018
	Healthy	0.977±0.010	0.954±0.018
	One-shot	0.977±0.010	0.954±0.018
U-Net	Full	0.982±0.007	0.964±0.012
	Healthy	0.93±0.09	0.89±0.13
	One-shot	0.961±0.021	0.93±0.04

came last, with the one trained on the Healthy dataset, Fig. 13e, having a Dice score of 0.637, and the model trained on the One-Shot dataset,



(a) Source. (b) Ground truth.

(c) Our pipeline. (d) U-Net FD.

(e) U-Net HD. (f) U-Net OSD.

Fig. 13: Comparison of symptomatic *Pst* predicted segmentation masks with the tested methods. Displaying results for one source image (a) with the corresponding ground truth (b). Predicted masks are shown for our method (c), and the U-Net models trained on the Full Training Dataset (d), the Healthy Dataset (e), and the One-shot Dataset (f).

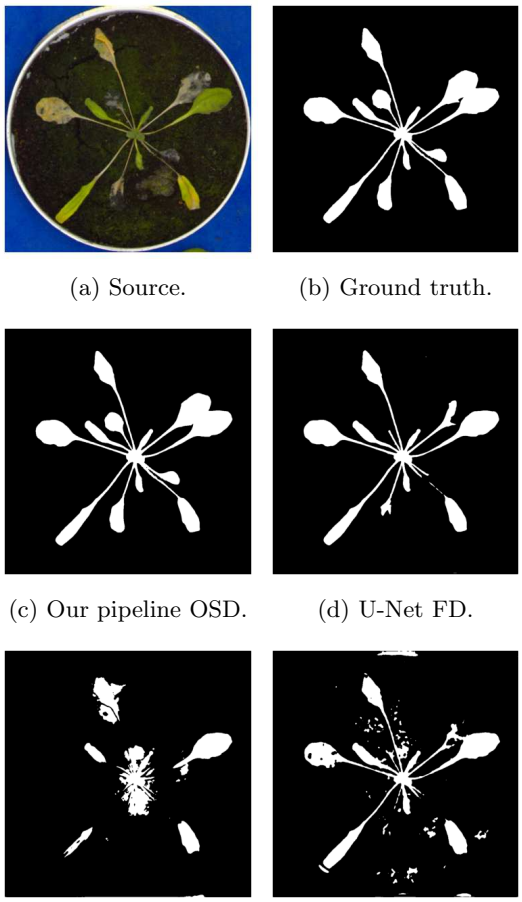
Fig. 13f, 0.757. Both models have severe over- and under-segmentation issues, with surprisingly better results for the model trained on a smaller dataset.

Finally, we compared the same models without further fine-tuning on the *Rps* Segmentation Test dataset. The OneRosette pipelines obtained the best scores with a drop of 0.01 in Dice and 0.027 in IoU compared to the *Pst* scores; all the pipelines obtained identical Dice scores. The U-Net model trained on the Full Training dataset came fourth with a drop of 0.023 in Dice and 0.038 in IoU. The remaining U-Net models came last, with more significant drops in both metrics.

Table 9: Dice and IoU with standard deviation on the hold-out *Rps* Segmentation Test Dataset for the best models of each approach trained on each dataset.

Method	Dataset	Dice	IoU
OneRosette	Full	0.97±0.04	0.94±0.07
	Healthy	0.97±0.04	0.94±0.07
	One-shot	0.97±0.04	0.94±0.07
U-Net	Full	0.96±0.05	0.93±0.08
	Healthy	0.88±0.16	0.82±0.21
	One-shot	0.91±0.08	0.85±0.11

As we did with the *Pst* segmentation experiment, we used a symptomatic image from the last acquisition timestamp, Fig. 14a, to compare the predictions of our pipeline trained on the One-shot dataset to the U-Net models. The ground truth to which all predictions were compared is shown in Fig. 14b. Once again, our pipeline, Fig. 14c, made the best prediction with a Dice score of 0.979; no symptomatic leaves were missed. The U-Net model trained on the Full Dataset, Fig. 14d, obtained a Dice score of 0.85 with portions or whole symptomatic leaves missing in the predicted mask. The U-Net model trained on the One-shot dataset, Fig. 14f, scored 0.756 in the Dice metric with missing leaves and background portions



(a) Source. (b) Ground truth.

(c) Our pipeline OSD. (d) U-Net FD.

(e) U-Net HD. (f) U-Net OSD.

Fig. 14: Comparison of symptomatic *Rps* predicted segmentation masks with the tested methods. Displaying results for one source image (a) with the corresponding ground truth (b). Predicted masks are shown for our method (c), and the U-Net models trained on the Full Training Dataset (d), the Healthy Dataset (e), and the One-shot Dataset (f).

added to the predicted mask. Finally, the U-Net model trained on the Healthy Dataset, Fig. 14e, showed the worst results with a Dice score of 0.537.

4.2 Time Series Level Validation

P. syringae pathovar tomato strain DC3000 Qualitative Analysis

Table 10: Occurence count and type for the variables found when applying EQ. 9 to the boostrapped *Pst* Qualitative Train Dataset.

Variable	Occurrences	Variable type
Homogeneity	710	texture
ASM	208	texture
Energy	49	texture
HSV H mean	13	color
Lab b std	13	color
Dissimilarity	4	texture
Lab b mean	2	color
Lab L std	1	color

Table 10 shows the results from the variable occurrences when looking for the best-discriminating feature using the bootstrapped *Pst* Qualitative Train Dataset. The top three variables accounted for 96.7% of the occurrences. The top three variables all came from the GLCM. The remaining five variables came mostly from color features and represented 3.3% of the occurrences. No morphological variable was selected in any instance. The most often selected variable is *homogeneity*, with 71% of the occurrences.

Once the best variable was selected, we calculated independent ANOVAs with the *Pst* Qualitative Validation dataset from D21-12h, the last timestamp before infection, until D24-12h, the first timestamp at which all infected plants were symptomatic. We present the ANOVAs results in Tab. 11. No significant difference was found

Table 11: ANOVAs for each selected timestamp calculated using *homogeneity* as the variable and treatment and tray as interactions. $F-statistic$ and $p-values$ are displayed for each source of variance at each timestamp.

	$F-statistic$			$p-value$		
Source of variance	tray	tray:treatment	treatment	tray	tray:treatment	treatment
timestamp						
D21-12h	1.99	1.16	0.27	0.07	0.33	6.07E-01
D22-15h	2.80	0.93	38.03	0.01	0.49	1.83E-08
D23-12h	2.54	0.60	292.87	0.02	0.76	2.43E-30
D23-15h	2.55	0.99	390.93	0.02	0.44	6.91E-35
D24-12h	2.27	1.40	1159.41	0.04	0.21	6.15E-54

between infected and control plants before infection. The data from the first recorded timestamp after infection showed a significant difference in the treatment as a source of variance between the control and the infected plants with a $p-value$ of $1.83E-8$, which allowed us to reject the H_0 hypothesis on the first available timestamp after infection. After the infection timestamp, the ANOVAs show significant differences between the control and infected plants regarding their tray position. The $p-values$ range from 0.01 to 0.04, much lower than the ones for the treatment, and the $F-statistic$ is low at all timestamps; furthermore, the ANOVAs show no interaction between the plant's treatments and their tray position.

Ralstonia pseudosolanacearum

Quantitative Analysis

Next, we present the results for the *Ralstonia pseudosolanacearum* quantitative disease index. Tab. 12 shows the results when looking for the best-discriminating variable using Eq. 9 on the bootstrapped *Rps* Qualitative Train Dataset. All selected variables came from the color-extracted features, with Lab b mean being selected over 50% of the times and Lab b standard deviation coming second with over 40% of the occurrences. The ANOVAs, Fig. 15c, detected a significant difference between the control and infected plants at D31-13h with a p-value of $2.8E-03$ and an F-statistic of $9.4E$ when the average disease index was still below 1.0. The difference stayed significant, with p-values decreasing up to $2.9E-42$ and F-statistics increasing to 540 by the end of the experiment. The other sources of variation remain under the 0.5 threshold for the duration of the

Table 12: Occurence count and type for the variables found when applying EQ. 9 to the boostrapped *Rps* Qualitative Train Dataset.

variable	occurrences	var type
Lab b mean	508	color
Lab b std	420	color
YIQ I std	25	color
RGB R std	21	color
Lab a std	11	color
RGB R mean	8	color
HSV H mean	7	color

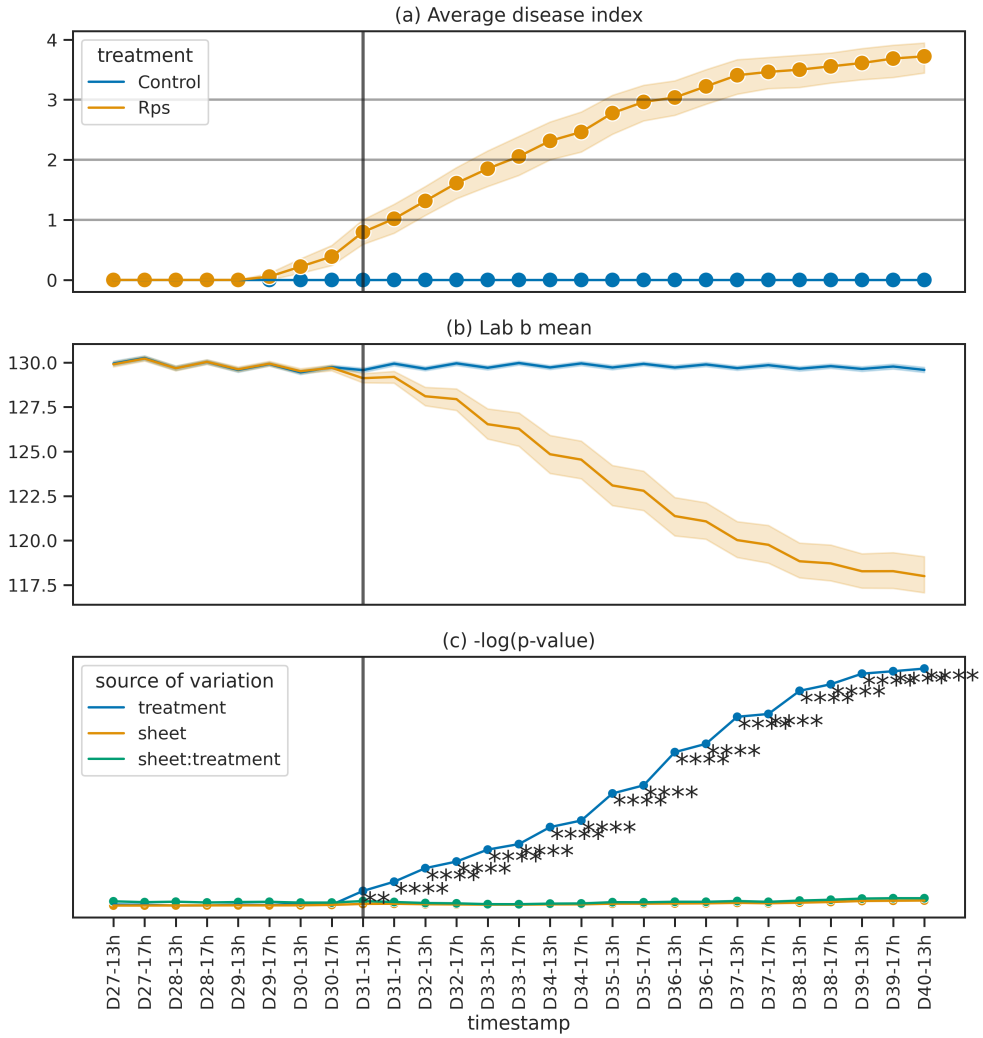


Fig. 15: From top to bottom: (a) average disease index over time. (b) Evolution of the selected variable, Lab b mean, over time for the control and the infected plants. (c) The ANOVAs $-log(p-value)$ for each source of variation over time. The vertical black bar at the timestamp indicates where the difference between control and infected plants is statistically significant.

experiment. The average disease index, Fig. 15a, and the Lab b mean variable, Fig. 15b, have a Pearson correlation score of -0.928 with a p-value lower than $2.2E-16$.

5 Discussion

In this paper, we proposed a novel one-shot method, "OneRosette", to segment *A. thaliana* plants' time series growing in pots. Our method

trained on a single asymptomatic image has similar metrics to a U-Net model trained on a full dataset containing symptomatic plants. It outperforms U-Net models trained without symptomatic plants. We tested the same OneRosette pipeline without further fine-tuning on a new pathosystem and obtained similarly good prediction scores, while the U-Net models obtained worse scores. This indicated that our one-shot learning approach could be invariant toward the selected pathosystems. Since we don't train a neural network, once the hyperparameters described in section 3.3 are found, setting up our pipeline takes far less time and fewer resources than training an entire U-Net neural network.

We used two experiments to assess whether the segmented plants' data extracted with OneRosette could be used in biological experiments. In the first case study, we used *A. thaliana* plants infected with *Pst* to assess the moment at which a significant difference in symptom presence could be detected between the control and infected plants with our segmentation method and compare the results to human annotations. Our method could discriminate between infected and control plants on the first available timestamp after infection when the expert annotated less than half of the plants as symptomatic.

In the second experiment, we used *A. thaliana* plants infected with *Rps*. This showed that our method can perform the same tasks with these two pathogens without further fine-tuning. Despite the new pathosystem, we obtained good segmentation scores (cf. Tab. 9) and found a statistical difference between the control and infected group at D31-13h, 44 hours after the expert annotated the first symptoms, while the average disease index was still under 1. Visually, the evolution of the disease index, Fig. 15a, and the selected variable, Lab b mean, Fig. 15b, appeared similar, which was confirmed by their Pearson correlation coefficient of 0.91.

This work could be expanded in multiple ways. In this study, we tested diseases caused by two bacterial agents. It would be interesting to test the pipeline with other pathogens, such as fungi like *Sclerotinia sclerotiorum* or a virus like the turnip mosaic virus; both pathogens can infect *A. thaliana* and develop different symptoms. It would also be interesting to extend this method to other plants like *Solanum lycopersicum*, also a host of both studied pathogens, could be interesting. This would allow for the better characterization of symptoms like wilting, which is better described by the plant's centroid vertical evolution [29], a variable that is challenging to estimate from a top view. Finally, in this study, we used images exclusively generated in a controlled conditions environment. It would be interesting to measure how using a time series of plants grown in the field would impact the method's performance.

6 Conclusion

A. thaliana is one of the leading plant models used to study plant-pathogen interactions. The segmentation of symptomatic whole plants as part of high-throughput phenotyping is a significant challenge requiring either training deep learning models for each pathosystem or limiting the study to detached leaves. This paper introduced a one-shot learning pipeline leveraging automatic prompting on a visual foundation model capable of segmenting time series of symptomatic *A. thaliana* plants infected with one of two bacterial pathogens. We obtained a Dice score of 0.977 with plants infected with *Pst* and 0.966 when infected with *Rps*. We successfully used our method to perform quantitative and qualitative analysis to describe plant-pathogen interactions. We believe that the low annotation requirements and the method's ease of use will make it easy to transfer to other pathosystems.

Supplementary information.

Acknowledgments. Fabrice Devoiles and Mehdi Khafif for assisting in plant production, maintenance, and growth in the automated greenhouse. This PhD thesis was accredited by INRAE Metaprogramme DIGIT-BIO.

Declarations

Author's contributions. FMM, SWM, DR, and NP conceived and designed this work. NP managed the plant material and the infections. FMM and NP carried out image acquisitions. FMM and NP carried out image annotations. FMM performed the statistical analysis. FMM and DR conceived the image processing algorithm and associated annotation tools. FMM, SWM, DR, and NP wrote and revised the manuscript. SWM, DR, and NP supervised the work. All authors read and approved the final manuscript.

Funding. The French National Infrastructure on Plant Phenotyping, ANR PHENOME 11-INBS-0012, funded this work and Felicià Maviane Maciá's Ph.D..

Availability of data and materials. The data, models, and computer user interface used for leaf *A. thaliana* annotations will be available after acceptance upon reasonable request.

Ethics approval and consent to participate. Not applicable.

Consent for publication. Not applicable.

Competing interests. The authors declare that they have no competing interests.

Appendix A Annotation User Interface

We provided the expert with an online user interface to annotate *Pst* qualitative data and *Rps* quantitative disease index.

Sidebar

The file manager allows the user to download a clean annotation CSV file; then, the user can either upload an empty CSV or continue annotating an existing one.

The "treatment" combo box allows the selection of a treatment in case the experiment contains multiple pathogens. The "plant combo box" provides for the individual selection of plants, and underneath the "Show completed plants" checkbox, it toggles the display of annotated plants.

Main Window

The currently annotated image is displayed at the top. Overlay controls allow the user to zoom in and out or pan the image as needed. The controls below the image are from left to right:

- The discrete player control allows the user to navigate within the image time series.
- The disease index spin edit allows the modification of the score value of the current image. If the value is modified, all pictures after the current image will get the same score; this allows the user to annotate only changes instead of every One-Shot.
- An indicator that turns green once the plant has been annotated.
- The download button allows downloading a CSV file with the current annotation state. The downloaded file can be used as a starting point if annotation has to be done over multiple sessions.

References

[1] Yaschenko, A.E., Alonso, J.M., Stepanova, A.N.: Arabidopsis as a model for translational research. The Plant Cell, 065 (2024) https://doi.org/10.1093/plcell/koae065 . Accessed 2024-11-18

[2] Arvidsson, S., Pérez-Rodríguez, P., Mueller-Roeber, B.: A growth phenotyping pipeline for Arabidopsis thaliana integrating image analysis and rosette area modeling for robust quantification of genotype effects. New Phytologist **191**(3), 895–907 (2011) https://doi.org/10.1111/j.1469-8137.2011.03756.x . _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1469-8137.2011.03756.x. Accessed 2025-01-30

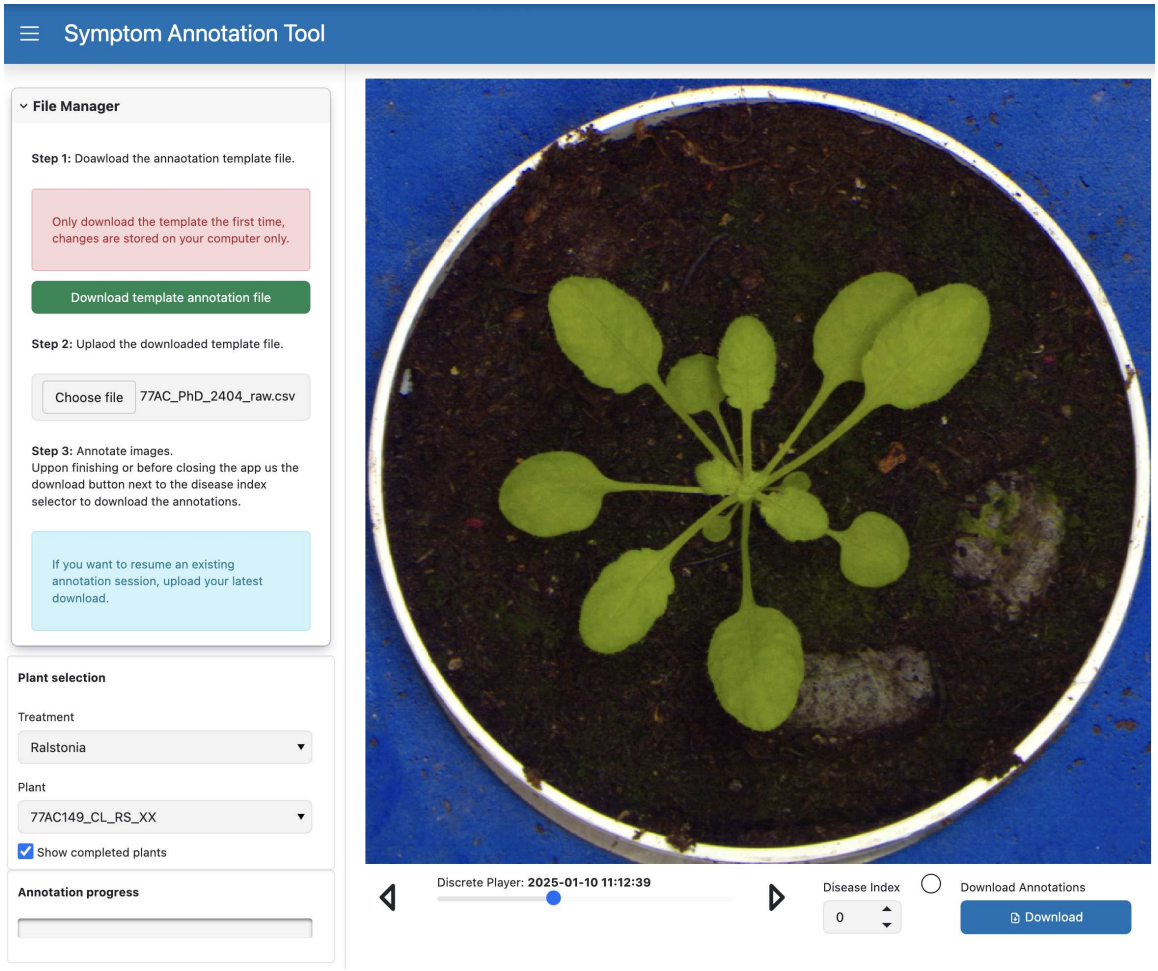


Fig. A1: User interface used to annotate *Pst* symptomatic data and *Rps* disease index.

[3] Whalen, M.C., Innes, R.W., Bent, A.F., Staskawicz, B.J.: Identification of Pseudomonas syringae Pathogens of Arabidopsis and a Bacterial Locus Determining Avirulence on Both Arabidopsis and Soybean. The Plant Cell **3**(1), 49–59 (1991) https://doi.org/10.2307/3869199 . Publisher: American Society of Plant Biologists (ASPB). Accessed 2025-02-12

[4] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L.: ImageNet: A Large-Scale Hierarchical Image Database https://doi.org/10.1109/CVPR.2009.5206848

[5] Yamada, K., Osakabe, Y., Mizoi, J., Nakashima, K., Fujita, Y., Shinozaki, K., Yamaguchi-Shinozaki, K.: Functional Analysis of an Arabidopsis thaliana Abiotic Stress-inducible Facilitated Diffusion Transporter for Monosaccharides. The Journal of biological chemistry **285**, 1138–46 (2009)

https://doi.org/10.1074/jbc.M109.054288

[6] Gardan, L., Shafik, H., Belouin, S., Broch, R., Grimont, F., Grimont, P.A.: DNA relatedness among the pathovars of Pseudomonas syringae and description of Pseudomonas tremae sp. nov. and Pseudomonas cannabina sp. nov. (ex Sutic and Dowson 1959). International Journal of Systematic Bacteriology **49 Pt 2**, 469–478 (1999) https://doi.org/10.1099/00207713-49-2-469

[7] Xin, X.-F., He, S.Y.: Pseudomonas syringae pv. tomato DC3000: A Model Pathogen for Probing Disease Susceptibility and Hormone Signaling in Plants. Annual Review of Phytopathology **51**(1), 473–498 (2013) https://doi.org/10.1146/annurev-phyto-082712-102321 . _eprint: https://doi.org/10.1146/annurev-phyto-082712-102321. Accessed 2023-04-24

[8] Pavicic, M., Overmyer, K., Rehman, A., Jones, P., Jacobson, D., Himanen, K.: Image-Based Methods to Score Fungal Pathogen Symptom Progression and Severity in Excised Arabidopsis Leaves. Plants **10**, 158 (2021) https://doi.org/10.3390/plants10010158

[9] Paauw, M., Hardeman, G., Taks, N.W., Lambalk, L., Berg, J.A., Pfeilmeier, S., Burg, H.A.: ScAnalyzer: an image processing tool to monitor plant disease symptoms and pathogen spread in Arabidopsis thaliana leaves. Plant Methods **20**(1), 80 (2024) https://doi.org/10.1186/s13007-024-01213-3 . Accessed 2024-06-04

[10] Lee, J.H., Lee, U., Yoo, J.H., Lee, T.S., Jung, J.H., Kim, H.S.: AraDQ: an automated digital phenotyping software for quantifying disease symptoms of flood-inoculated Arabidopsis seedlings. Plant Methods **20**(1), 44 (2024) https://doi.org/10.1186/s13007-024-01171-w . Accessed 2024-05-22

[11] Bochkovskiy, A., Wang, C.-Y., Liao, H.-Y.M.: YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv. Version Number: 1 (2020). https://doi.org/10.48550/ARXIV.2004.10934 . https://arxiv.org/abs/2004.10934 Accessed 2024-12-02

[12] Bell, J., Dee, H.M.: Aberystwyth Leaf Evaluation Dataset. Zenodo (2016). https://doi.org/10.5281/zenodo.168158 . https://zenodo.org/records/168158 Accessed 2024-07-10

[13] Minervini, M., Fischbach, A., Scharr, H., Tsaftaris, S.A.: Finely-grained annotated

datasets for image-based plant phenotyping. Pattern Recognition Letters **81**, 80–89 (2016) https://doi.org/10.1016/j.patrec.2015.10.013 . Accessed 2024-07-10

[14] Scharr, H., Minervini, M., Fischbach, A., Tsaftaris, S.A.: Annotated Image Datasets of Rosette Plants

[15] Deb, M., Garai, A., Das, A., Dhal, K.G.: LS-Net: a convolutional neural network for leaf segmentation of rosette plants. Neural Computing and Applications **34**(21), 18511–18524 (2022) https://doi.org/10.1007/s00521-022-07479-9 . Accessed 2024-07-03

[16] Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv. arXiv:1505.04597 [cs] (2015). https://doi.org/10.48550/arXiv.1505.04597 . http://arxiv.org/abs/1505.04597 Accessed 2023-05-08

[17] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.-C.: MobileNetV2: Inverted Residuals and Linear Bottlenecks. arXiv. arXiv:1801.04381 (2019). https://doi.org/10.48550/arXiv.1801.04381 . http://arxiv.org/abs/1801.04381 Accessed 2024-12-02

[18] Granier, C., Aguirrezabal, L., Chenu, K., Cookson, S.J., Dauzat, M., Hamard, P., Thioux, J.-J., Rolland, G., Bouchier-Combaud, S., Lebaudy, A., Muller, B., Simonneau, T., Tardieu, F.: PHENOPSIS, an automated platform for reproducible phenotyping of plant responses to soil water deficit in Arabidopsis thaliana permitted the identification of an accession with low sensitivity to soil water deficit. New Phytologist **169**(3), 623–635 (2006) https://doi.org/10.1111/j.1469-8137.2005.01609.x . _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1469-8137.2005.01609.x. Accessed 2024-12-02

[19] Katagiri, F., Thilmony, R., He, S.Y.: The Arabidopsis Thaliana-Pseudomonas Syringae Interaction. The Arabidopsis Book **2002**(1) (2002) https://doi.org/10.1199/tab.0039 . Publisher: The American Society of Plant Biologists. Accessed 2023-05-10

[20] Duda, R.O., Hart, P.E.: Use of the Hough transformation to detect lines and curves in pictures. Communications of the ACM **15**(1), 11–15 (1972) https://doi.org/10.1145/361237.361242 . Accessed 2023-12-12

[21] Bradski, G.: The OpenCV Library. Dr. Dobb's Journal of Software Tools (2000)

[22] Canny, J.: A Computational Approach To Edge Detection. Pattern Analysis and Machine Intelligence, IEEE Transactions on

PAMI-8, 679–698 (1986) https://doi.org/10.1109/TPAMI.1986.4767851

[23] Ramanath, R., Drew, M.S.: Color Spaces. In: Ikeuchi, K. (ed.) Computer Vision: A Reference Guide, pp. 123–132. Springer, Boston, MA (2014). https://doi.org/10.1007/978-0-387-31439-6_452 . https://doi.org/10.1007/978-0-387-31439-6_452 Accessed 2025-01-06

[24] Ravi, N., Gabeur, V., Hu, Y.-T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., Mintun, E., Pan, J., Alwala, K.V., Carion, N., Wu, C.-Y., Girshick, R., Dollár, P., Feichtenhofer, C.: SAM 2: Segment Anything in Images and Videos. arXiv. arXiv:2408.00714 [cs] (2024). http://arxiv.org/abs/2408.00714 Accessed 2024-08-04

[25] Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S. (eds.): SLIC Superpixels, EPFL (2010)

[26] Garcia-Castellanos, D., Lombardo, U.: Poles of inaccessibility: A calculation algorithm for the remotest places on earth. Scottish Geographical Journal **123**(3), 227–233 (2007) https://doi.org/10.1080/14702540801897809 . Accessed 2025-01-23

[27] Agafonkin, V.: Polylabel: a fast algorithm for finding the pole of inaccessibility of a polygon. original-date: 2016-07-08T16:34:29Z (2016). https://github.com/mapbox/polylabel Accessed 2025-01-23

[28] Skalski, P.: Make Sense (2019). https://github.com/SkalskiP/make-sense

[29] Gerlin, L., Escourrou, A., Cassan, C., Maviane Macia, F., Peeters, N., Genin, S., Baroukh, C.: Unravelling physiological signatures of tomato bacterial wilt and xylem metabolites exploited by. Environmental Microbiology **23**(10), 5962–5978 (2021) https://doi.org/10.1111/1462-2920.15535 . _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1462-2920.15535. Accessed 2025-01-06

[30] Iakubovskii, P.: Segmentation Models Pytorch (2019). https://github.com/qubvel/segmentation_models.pytorch

[31] Formin, V., Anmol, J., Desroziers, S., Kriss, J., Tejani, A.: High-level library to help with training neural networks in PyTorch (2020). https://github.com/pytorch/ignite

[32] Buslaev, A., Iglovikov, V.I., Khvedchenya, E., Parinov, A., Druzhinin, M., Kalinin, A.A.: Albumentations: Fast and Flexible Image Augmentations. Information **11**(2), 125

(2020) https://doi.org/10.3390/info11020125 . Number: 2 Publisher: Multidisciplinary Digital Publishing Institute. Accessed 2023-12-12

[33] Haralick, R.M., Shanmugam, K., Dinstein, I.: Textural Features for Image Classification. IEEE Transactions on Systems, Man, and Cybernetics **SMC-3**(6), 610–621 (1973) https://doi.org/10.1109/TSMC.1973.4309314 . Accessed 2024-12-23