

PHENOSCOPE: an automated large-scale phenotyping platform offering high spatial homogeneity.

Sébastien Tisné<sup>1,2</sup>, Yann Serrand<sup>1,2</sup>, Liên Bach<sup>1,2</sup>, Elodie Gilbert<sup>1,2</sup>, Rachid Ben Ameer<sup>1,2</sup>, Hervé Balasse<sup>1,2</sup>, Roger Voisin<sup>1,2</sup>, David Bouchez<sup>1,2</sup>, Mylène Durand-Tardif<sup>1,2</sup>, Philippe Guerche<sup>1,2</sup>, Gaël Chareyron<sup>3</sup>, Jérôme Da Rugna<sup>3</sup>, Christine Camilleri<sup>1,2</sup>, Olivier Loudet<sup>1,2,\*</sup>

## SUPPORTING INFORMATION

### **Figure S1.**

Diagram of pot movements on the Phenoscope table.

Pots (alternating gray and white squares) are represented at their initial position in which weighing, watering and imaging occurs (a). From this position, the 'one-step' displacement included in every single cycle is composed of two moves. First, the 39 connecting rods (green elements) powered by motors M2 and M4 (motors are under the table), and the left pusher arm (yellow element motored by motor M6) ensure a sequential ascending/descending move into a temporary position (b). Then, two rakes (red elements) motored by engines M1 and M5 and the return pusher arm (yellow element powered by motor M3) ensure an opposing horizontal move back into the initial position+1 for each pot (c); follow the example of the blue pot in the bottom left corner and watch <http://www.inra.fr/vast/Files/PhenoFilm.avi> for a visual, continuous demonstration of pot moves.

### **Figure S2.**

Flow chart of different modes used for Phenoscope experiment management.

The 'Cycle' box refers to the unitary cycle described in Figure 2.

### **Figure S3.**

System architecture of the Phenoscope platform.

### **Figure S4.**

Processing the Phenoscope images.

- (a) General processing scheme
- (b) Algorithm outline for the pre-processing step
- (c) Algorithm outline for the seed extraction step

**Figure S5.**

Example of phenotypic data extraction from a Phenoscope experiment.

Evolution of average color (a) and compactness (b) of Col-0 and Cvi-0 genotypes through time (21 days of growth) on Phenoscope. Compactness was obtained by dividing the projected rosette area by the area of the disk delineating the rosette.

**Figure S6.**

Physical representation of the Phenoscope database model.

**Figure S7.**

Cycling the pots on Phenoscope reduces heterogeneity.

Distribution of 24 h water loss for 735 plugs placed on the Phenoscope table when pots were cycling (solid line) or at fixed positions (dotted line). Bar plots in the inset indicate the coefficient of variation (CV) among 24 h water loss when the plugs were cycling (black bar) or at fixed positions (striped bar). This figure is from the same dataset as that represented in Figure 4.

**Figure S8.**

Phenotypic variation among Cvi-0 x Col-0 RILs.

Frequency distributions of the average final projected rosette area (cm<sup>2</sup>) and relative expansion rate in the Cvi-0 x Col-0 population grown under control (60% SWC, blue bars) and water deficit (30% SWC, orange bars) treatments. Mean values for parental lines Cvi-0 (gray arrows) and Col-0 (black arrows) are indicated over each distribution (n=8).

**Figure S9.**

Comparing QTL analysis from two independent biological replicates.

QTL analysis for final projected rosette area in the Cvi-0 x Col-0 RIL set grown under control (blue) and water deficit (orange) treatments. QTL mapping was performed using individual phenotypic values for each RIL from two independent experiments (solid and dotted lines). LOD scores from 'Multiple QTL Mapping' are plotted along the five chromosomes, the sign indicates

the direction of the allelic effect as in Figure 7. Significance thresholds (5%) determined by a permutation test are plotted as horizontal dotted lines.

**Figure S10.**

Phenotypes obtained with HIFs designed to confirm QTLs located on chromosome 2 and 4 as indicated in Figure 7b.

Mean and standard error of the final projected rosette area (cm<sup>2</sup>), the relative expansion rate and the projected rosette area 11 days after sowing (11 DAS PRA; cm<sup>2</sup>) are plotted for individuals carrying either Col (black) or Cvi (gray) alleles at the segregating region (see Figure 7b) from the indicated RIL (a-l). When significant, the allelic effect from plants grown under control (60% SWC) or water deficit (30% SWC) treatments and the allele x treatment interaction is indicated by \*, \*\*, and \*\*\* (respectively indicating significant differences with P values <0.05, <0.01 and <0.001).

**Table S1.**

Parameters of the significant QTLs detected for growth traits in the Cvi-0 x Col-0 RIL set. PRA, Projected Rosette Area (cm<sup>2</sup>); RER, Relative Expansion Rate.

Trait	Treatment	$r^2G^a$	$r^2P^b$	Marker <sup>c</sup>	Position (cM)	$F$	P-value <sup>d</sup>	$r^2P^e$	$2a^f$
Final PRA	Control	0.74	0.61	c1_02212	6.0	103.76	<2.2e-16	0.16	1.26
				c1_05593	24.6	21.75	4.997e-06	0.03	1.18
				c1_12295	54.4	11.22	0.000926	0.02	-0.68
				c1_28667	130.6	18.81	2.071e-05	0.03	0.76
				c2_10250	44.5	78.02	<2.2e-16	0.12	-1.29
				c2_17606	80.1	100.03	<2.2e-16	0.15	-1.91
				c4_10609	51.3	36.94	4.391e-09	0.06	1.25
				c5_24997	106.5	37.03	4.212e-09	0.06	1.14
Final PRA	Water deficit	0.73	0.58	c1_02212	6.0	69.14	5.170e-15	0.11	0.76
				c1_05593	24.6	25.44	8.543e-07	0.04	0.61
				c1_28667	130.6	9.12	0.002766	0.02	0.33
				c2_10250	44.6	69.42	4.624e-15	0.11	-0.87
				c2_18753	86.9	159.70	<2.2e-16	0.25	-1.88
				c4_11878	56.8	8.69	0.003480	0.01	0.47
				c5_24997	106.5	20.68	8.315e-06	0.03	0.66
				RER	Control	-	0.42	c1_28454	129.6
c2_16837	77.4	87.85	<2.2e-16					0.21	-0.0109
c3_01901	4.7	13.84	0.000246					0.03	0.00544
c3_06631	24.2	6.21	0.013331					0.02	-0.00301
c4_14819	68.8	7.81	0.005597					0.02	-0.00275
c5_19316	77.7	26.00	6.770e-07					0.06	0.00426
c5_26671	112.8	24.53	1.354e-06					0.06	0.00636
RER	Water deficit	-	0.47					c1_28454	129.6
				c2_18753	86.9	192.98	<2.2e-16	0.35	-0.01525
				c4_11878	56.8	16.17	7.383e-05	0.03	-0.00384
				c5_14766	58.5	18.56	2.257e-05	0.03	0.00372
				c5_24997	106.5	24.92	1.035e-06	0.05	0.00578

<sup>a</sup> Percentage of genotypic variance explained by the QTL model.

<sup>b</sup> Percentage of phenotypic variance explained by the QTL model.

<sup>c</sup> Marker determined in MQM cofactor selection (see Experimental procedures).

<sup>d</sup> Significance of the term of the QTL model (see Experimental procedures).

<sup>e</sup> Percentage of phenotypic variance explained by each term of the QTL model.

<sup>f</sup> Mean effect of the replacement of both Col alleles by Cvi alleles at the QTL.

## SUPPORTING EXPERIMENTAL PROCEDURES

**Methods S1.** Detailed technical description of the Phenoscope set-up.

### - The Phenoscope table

A Phenoscope table holds 735 individual plastic pots (5 x 5 cm) maintained in a closed-circuit track by rails. Pot movements are ensured by 'connecting rods', 'rakes' and 'pushers arms' powered by six independent motors (M1 to M6; AC induction motors). They are coupled with a parallel shaft gear and are driven by an inverter whose speed can be set. For so-called 'vertical' displacements (ascending and descending movements, cf. Figure S1), connecting rods are used (20 for the top and 19 for the bottom) powered by M2 and M4. They are attached to a shaft that is moved by a rod connected to the output of the motor's shaft gear by a ball and socket. The left and return pusher arms work like the connecting rods and are powered by M6 and M3, respectively. Two rakes are used for 'horizontal' displacement. They are powered by motors M1 and M5, and linked to the motors by a rod. Each displacement unit is composed of the succession of two moves. First, the connecting rods and the left pusher are activated and push the pot to a temporary position. Then, the rakes and the return pusher allow the transversal movement to the initial position +1 for each pot (Figure S1). Weighing is performed by a platform leaning on a single point load cell (Scaime AR0.6). This cell is connected to a weighing indicator (Scaime IPE 50 P) that displays the value and transfers it via a RS232 link. A RS232/Ethernet converter is then used to transfer data to the programmable logic controller (PLC). Watering is performed by a robust peristaltic metering pump (Pompes AB8 SAN). Images are taken by a digital color matrix camera (Baumer TXG20c) at a resolution of 1624 x 1236 pixels. It is connected directly to the computer by a GigaEthernet cable.

### - The Phenoscope database

The Phenoscope DB uses PostgreSQL (<http://www.postgresql.org/>) as database management system (DBMS), a free, multi-platform and open-source software. It uses SQL language to manage the database structure and schema (DDL), to insert, delete or update data in tables and perform read-only queries (DML), to control access to the database (DCL), and finally to manage the changes made by DML statements (TCL). From a conceptual data model, where all the data requirements are initially recorded and organized into entities and relationships, a physical data model now organizes data into tables and columns. The Phenoscope DB structure and model is shown in Figure S6. The Phenoscope table on which the experiment is performed is described in

the 'Phenoscope' table. The characteristics of the growth room are stored in the 'GrowthRoom' table. The experiment itself is described in the 'Experiment' table that includes the user name, the title, start and end date and the description of the experiment. The 'PhenoUser', 'GroupUser' and 'PhenoGroup' tables manage users and groups and define their rights. For each experiment, a maximum of 735 pots can be used. Their empty (tare) weight (acquired from Mode #1), reference weight (acquired from Mode #2) and the sowing date are stored in the 'Pot' table. The plant identity is also stored in the 'Pot' table. It refers to the 'Plant' table, where the genotype and a description of the plant are detailed. Each element of the 'Pot' table is associated with an experiment and a scenario. The scenario defines the protocol of the experiment for one plug. A scenario is first defined by the number of days it lasts, and the number of cycles (and their duration) per day. These data are listed in the 'Cycle' table. Then each cycle is associated with an instruction that defines how Phenoscope will treat the plug. The 'Pump' table (like the 'Sensor' table) declares the pumps (sensors) and describes them; several pumps or sensors per Phenoscope can be used. The fields of the 'WateringActivation' table are the type of watering (in volume or in percentage of SWC) and its value combined with the pump used. The measured parameters are separated into two groups: environmental parameters recorded by probes in the growth room, and data records related to the pot/individual (pot weights, watering, phenotypic measures extracted by *Phenospeed*). New variables can be added at any time. Finally, the 'Tag' and 'TypeTag' tables identify outliers (individual plants) for/from the statistical analyses. A management interface is available for the user to supply the database with all necessary information prior to the start of an experiment to describe the scenario and the desired protocol.

#### **- The Phenoscope analysis suite**

The *Phenospeed* application includes all necessary tools and algorithms to extract estimates of rosette parameters. The main steps of the application are described in Figure S4a. The first step is to compute automatically, for each image and pot, the segmentation of the rosette and to measure its morphological properties. The second step relies on the expert user's analysis: only pots detected as atypical are examined. This allows the expert to validate or invalidate atypical results, for example when the plant is missing, when image segmentation is poor or when a leaf from another pot overlaps in the image. When the initial analysis is considered invalid, the user can recompute the pot's image analysis using new parameters (overlap threshold, bounding box). The main elements of *Phenospeed* are the image processing library, the processing interface and the atypical result-checking interface. The image processing library was developed

in C++ (please refer to the module 'CILAutoseed' developed within the library 'OpenCIL', which source code is available from our website: <http://labs.esilv.fr/3i/openCIL/>), except for the morphological measures. The well-known library 'openCV' (<http://opencv.org/>) is used to compute these measures, and to integrate features like eccentricity, compactness or texture descriptor. User interfaces are .NET/C# applications with user-friendly GUI dedicated to each task. The image processing core is multi-threaded — several pots are simultaneously computed — and provides a result in a reasonable time on a regular multi-core computer.

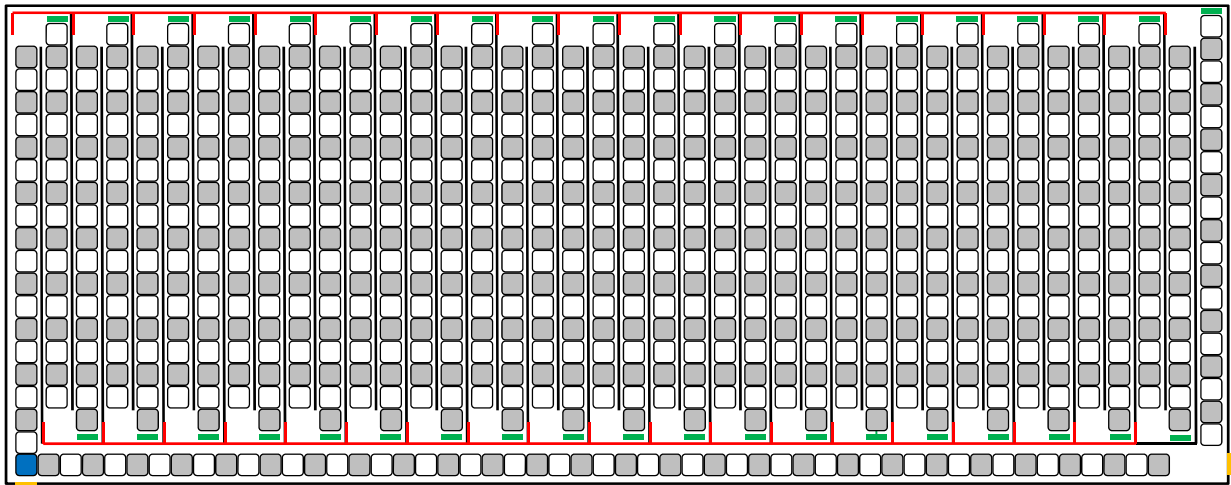
In more detail (Figure S4b), each image acquisition undergoes a pre-processing step, principally white balance color correction. This automatic white color balance is reached using scene illuminant estimation and classic von Kries method (Sharma, 2002; Tominaga and Wandell, 2002). Representative colors, the segmentation seeds, are then computed for each pot. Seed extraction (Figure S4c), representing colors of the rosette, is done using color histogram processing. The peak-finding algorithm introduced previously (Cheng *et al.*, 2001) is applied to the RGB color variation histogram. To enhance robustness of the detection against color variability, all detected peaks are gathered in the same set and a k-means algorithm separates the most significant seeds. Experiments using various lighting, camera and conditions prove that this approach is able to recover correctly robust seeds. The local region growing segmentation (Zhang *et al.*, 2008) is initiated using these seeds. Low-level morphological operations are applied at the end of the process to remove artifacts and close the shape. During this step, the majority of overlaps from neighboring pots are removed. Finally, rosette characterization is done through well-known parameter estimates: area (estimated in pixels and then calibrated in cm<sup>2</sup> with a scale), color mean and dispersion, bounding circle radius. Using predefined rules, the last process defines, for each image, metadata representing the segmentation assessment. For instance, if the rosette area experiences an unexpected decrease or if the segmented rosette is larger than the camera field-of-view, the image will be marked as atypical for subsequent treatment or discarding.

## REFERENCES

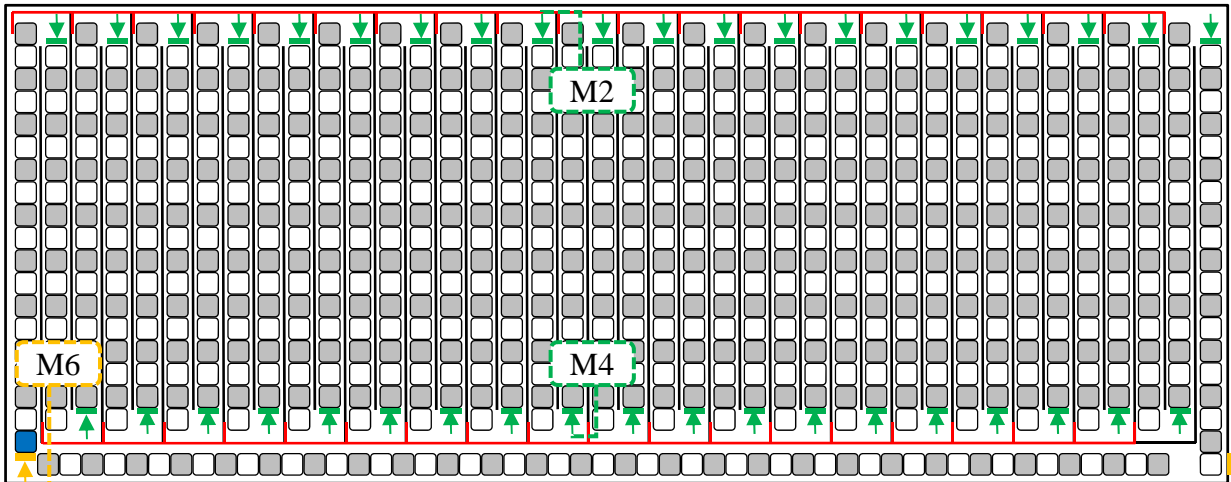
- Cheng, H., Jiang, X., Sun, Y. and Wang, J.** (2001) Color image segmentation: advances and prospects. *Pattern Recognit.* **34**, 2259-2281.
- Sharma, G.** (2002) Digital Color Imaging Handbook. CRC Press.
- Tominaga, S. and Wandell, B. A.** (2002) Natural scene-illuminant estimation using the sensor correlation. *Proceedings of the IEEE* **90**, 42-56.
- Zhang, H., Fritts, J. E. and Goldman, S. A.** (2008) Image segmentation evaluation: A survey of unsupervised methods. *Comput. Vis. Image Underst.* **110**, 260-280.

# Supplementary Figure S1

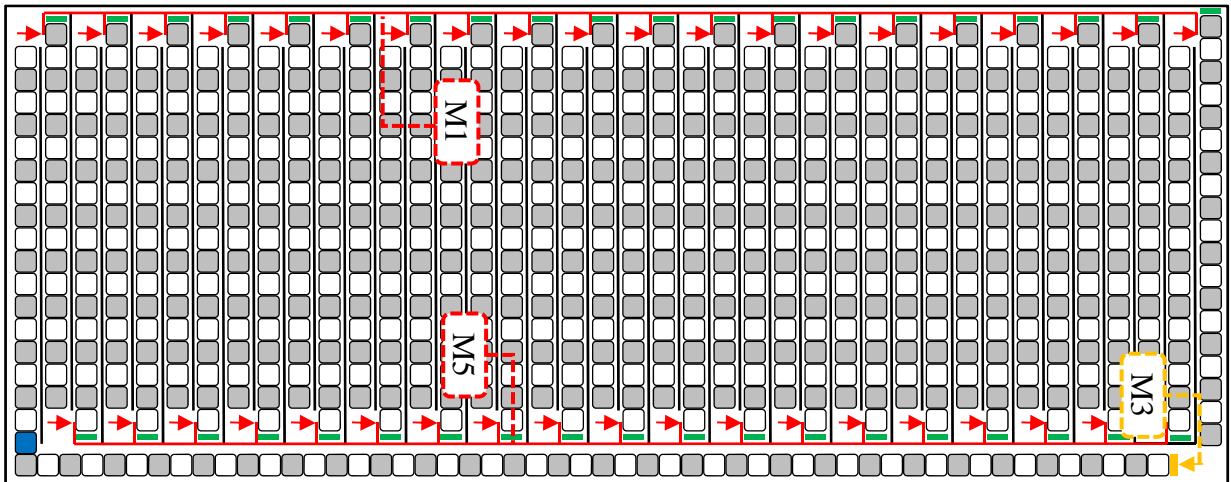
(a)



(b)



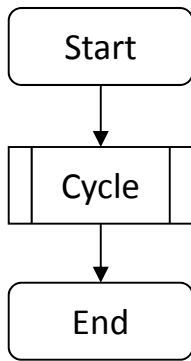
(c)



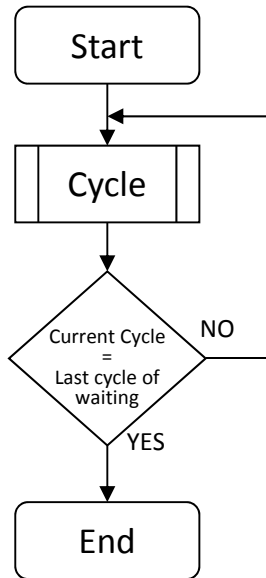


# Supplementary Figure S2

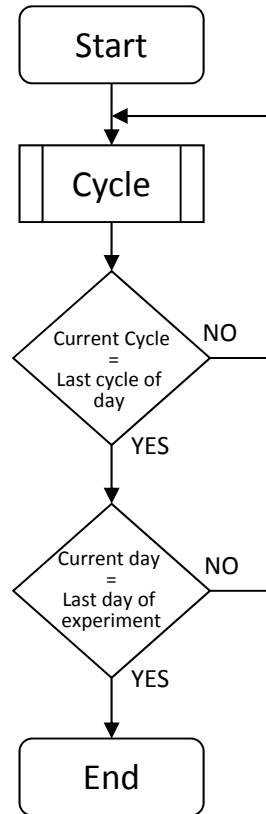
Mode #1 / Mode #2



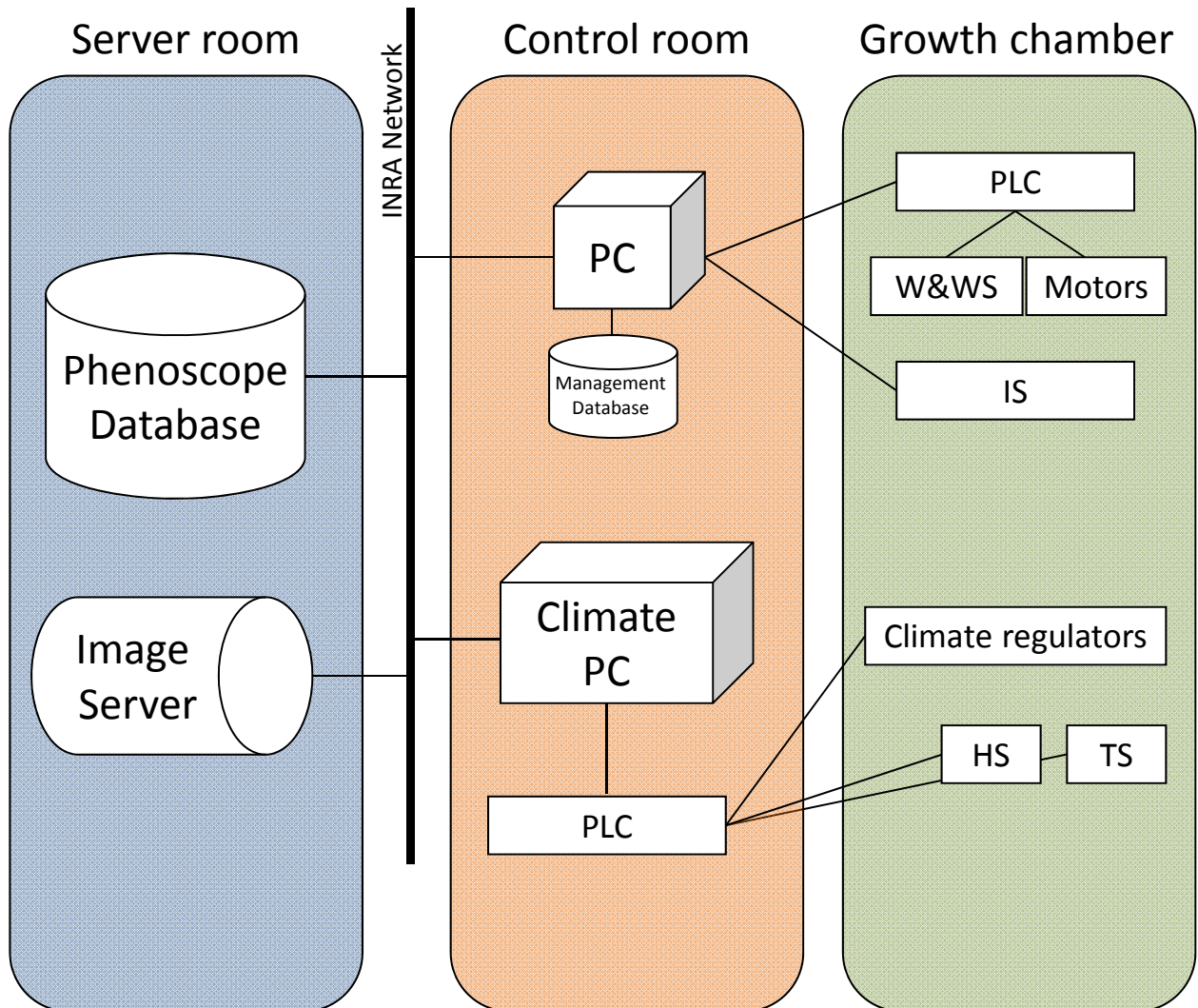
Mode #3



Mode #4



## Supplementary Figure S3



PLC : programmable logic controller

IS: imaging station

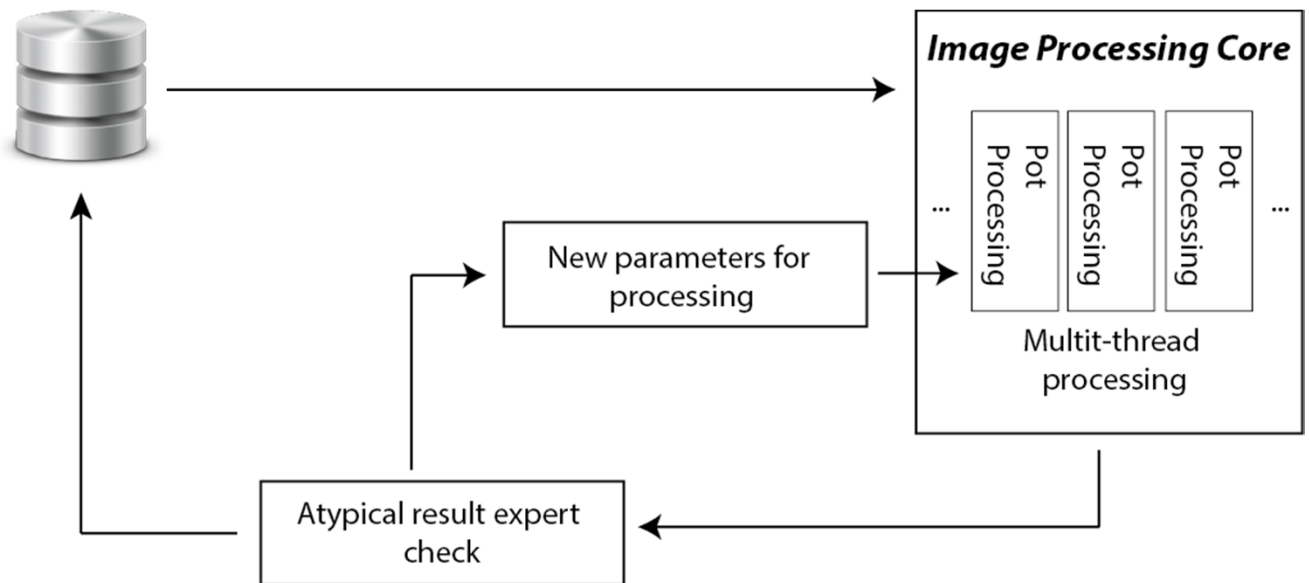
W&WS : weighing and watering station

HS : humidity sensor

TS : temperature sensor

# Supplementary Figure S4

(a)



(b)

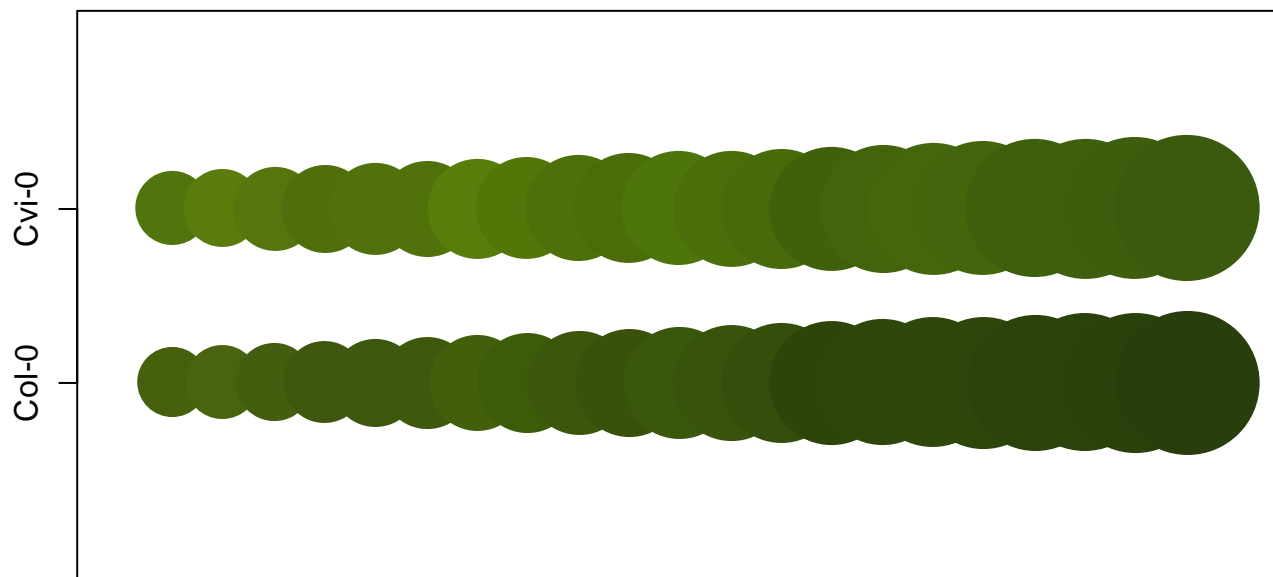
```
Input: All images of the experiment
Output: Morphological analysis
for each image do
    White balance preprocessing
end
for each pot do
    Seed extraction
end
for each pot do
    Seed-based threshold segmentation
    Local region growing segmentation
    Final morphological operations
end
Rosette characterization
Result analyzing and reporting
```

(c)

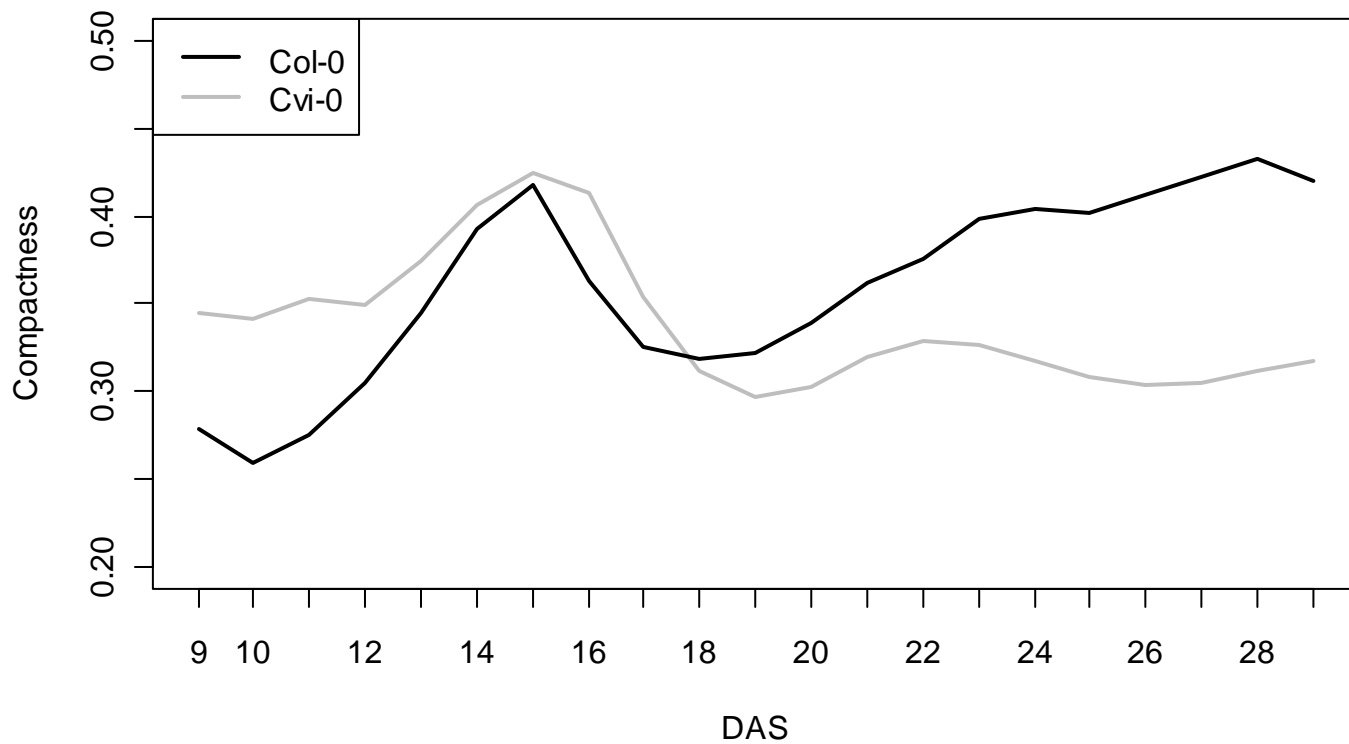
```
Input: Pot image sequence
Output: Color segmentation seeds
for each pot do
    Extract histogram variation along the sequence
    Extract significant seeds by peak detection
    Insert seeds into the pre-seeds set
end
K-Means on pre-seeds set
Most significant seed selection
```

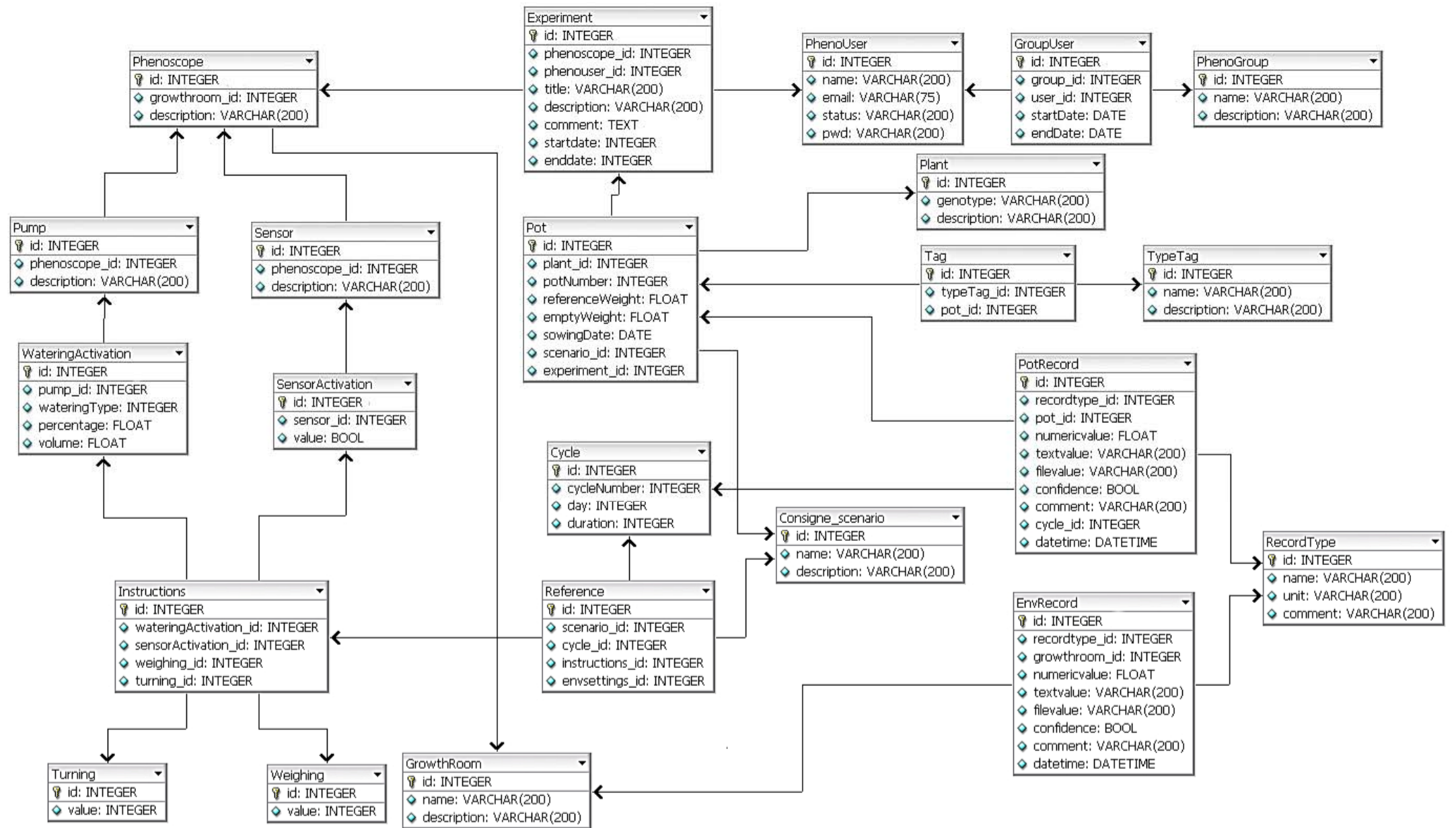
# Supplementary Figure S5

(a)

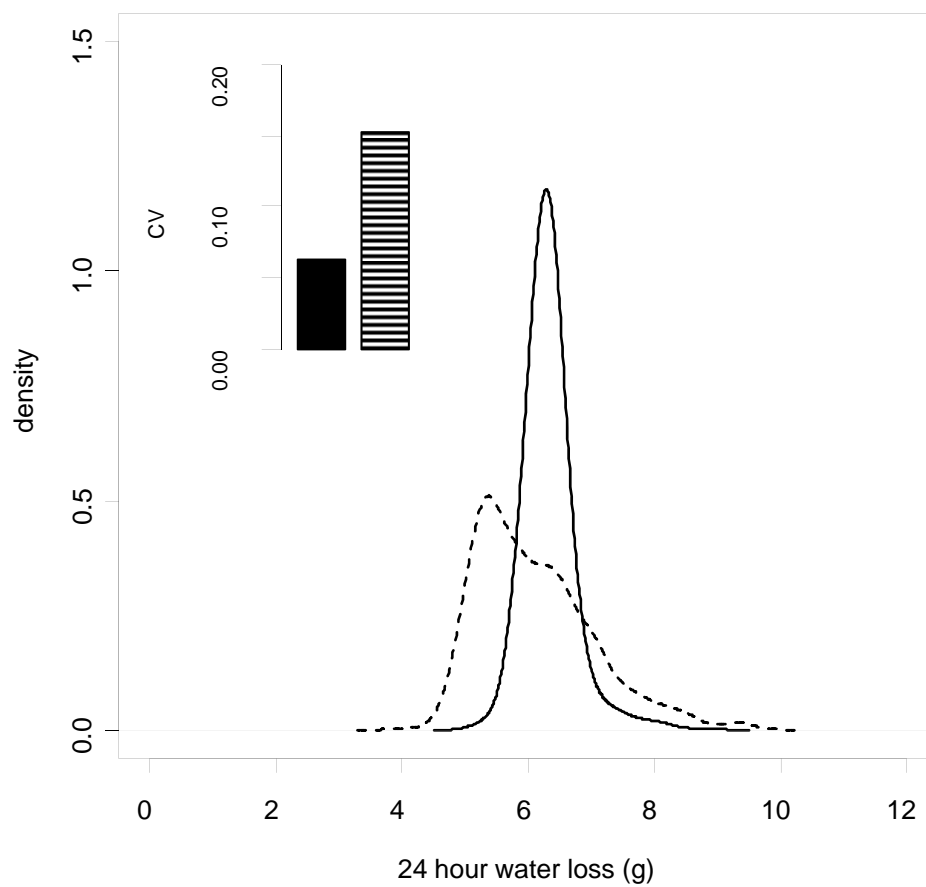


(b)

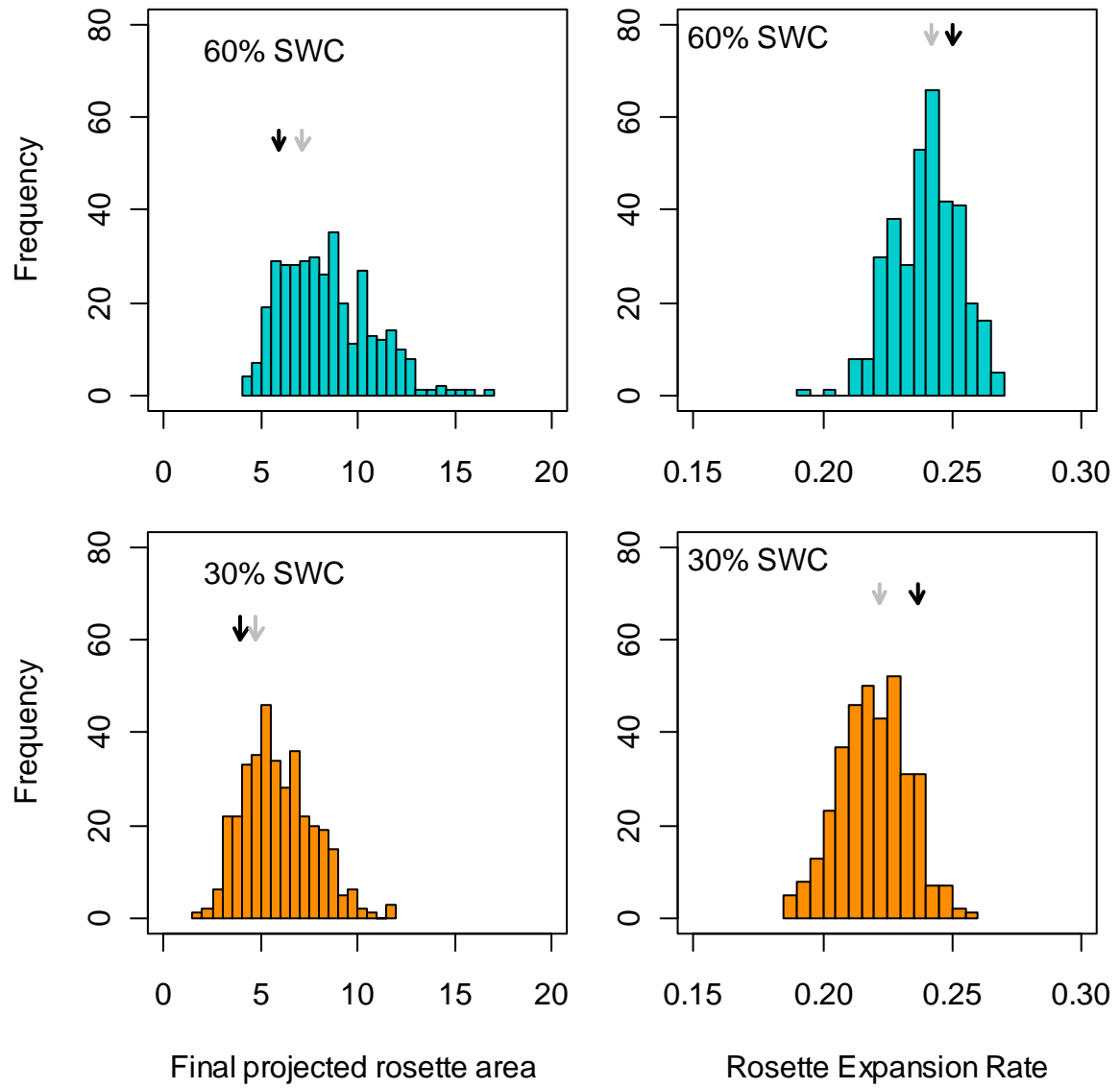




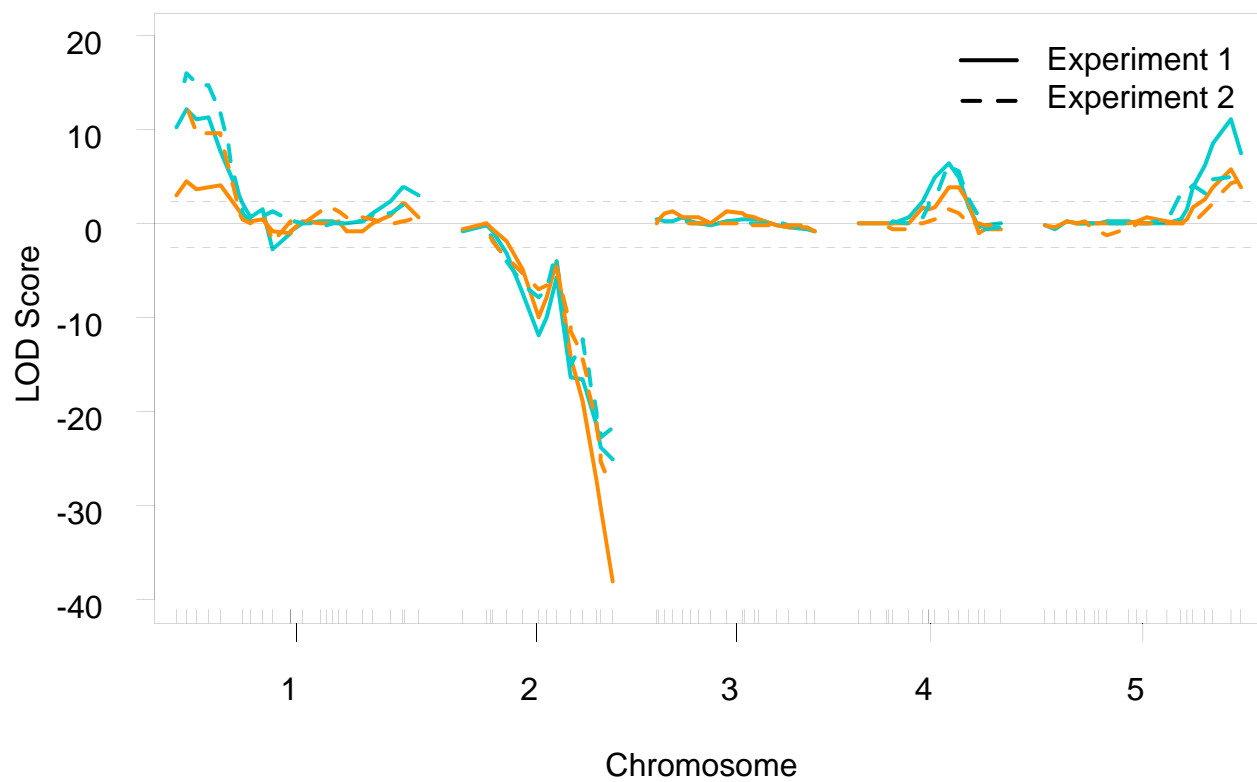
# Supplementary Figure S7



# Supplementary Figure S8



Supplementary Figure S9

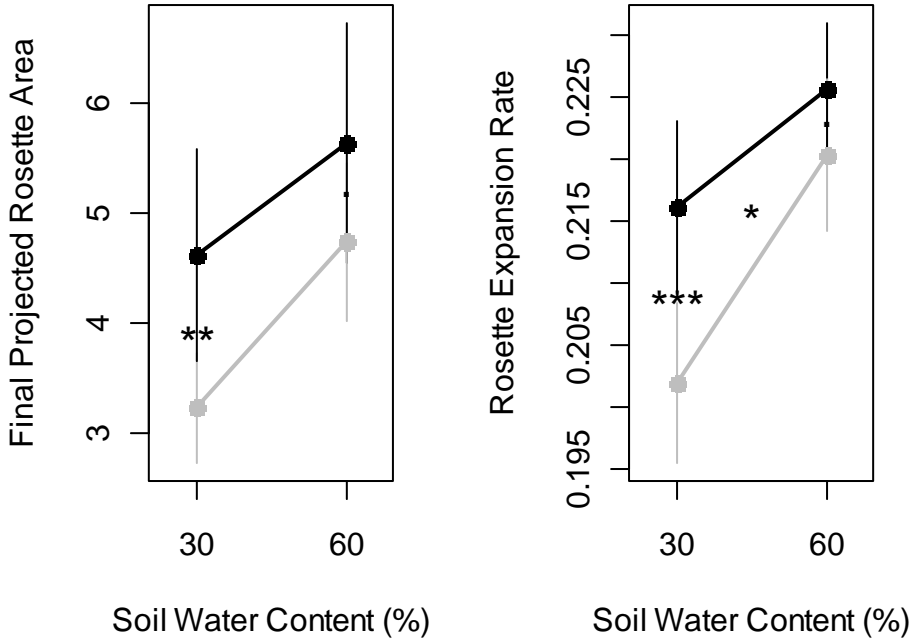




# Supplementary Figure S10

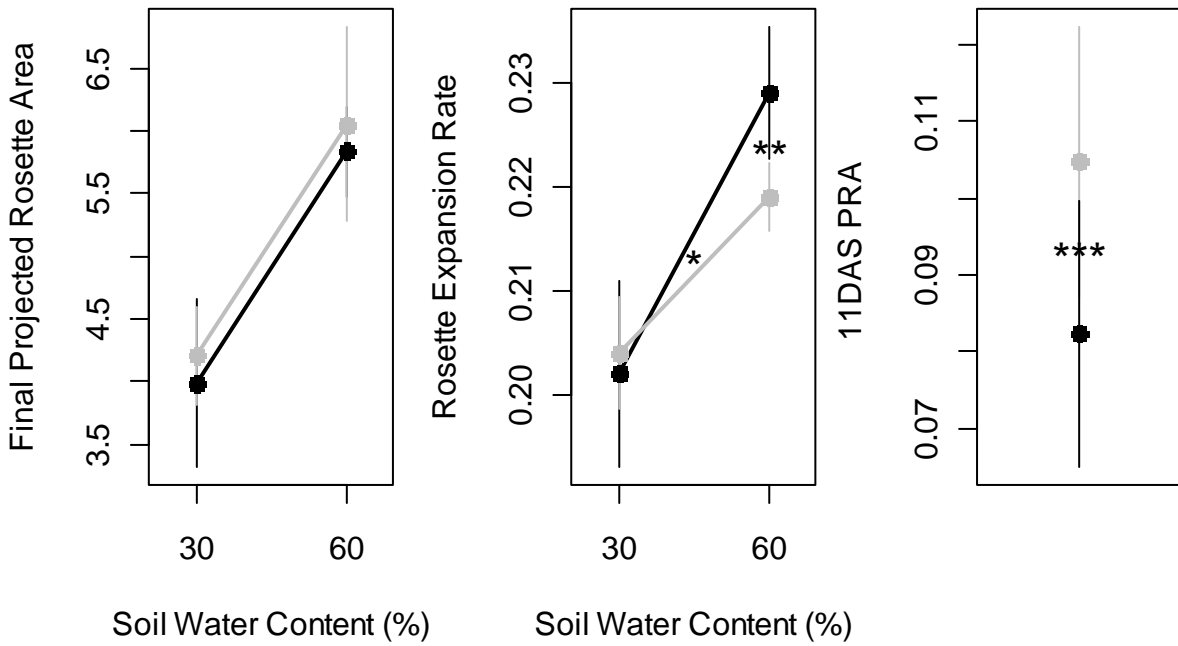
(a)

**8HV411**



(b)

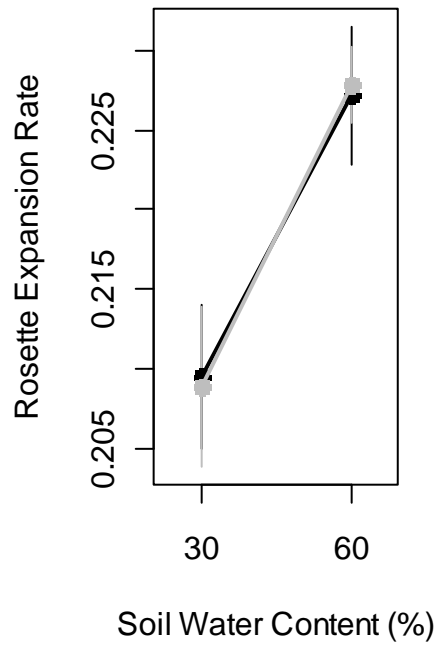
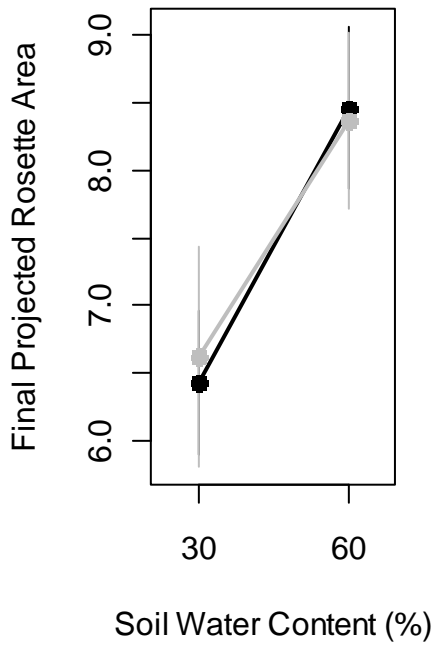
**8HV272**



# Supplementary Figure S10

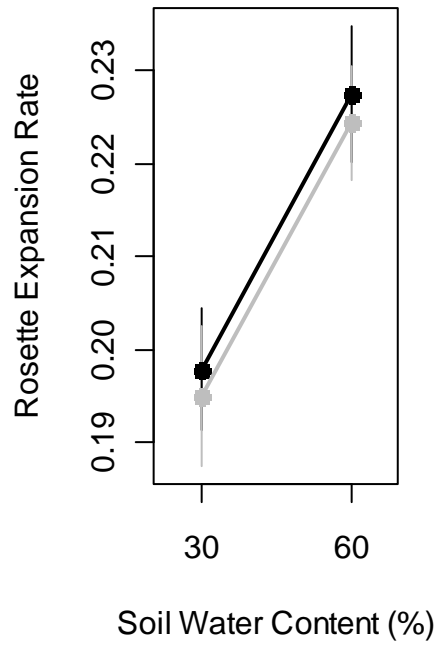
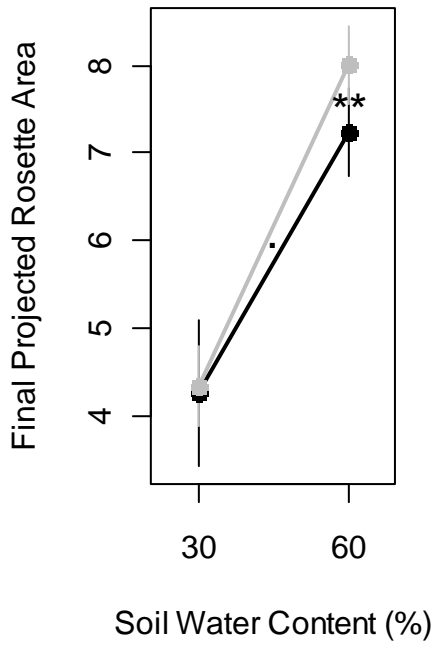
(c)

**8HV418**



(d)

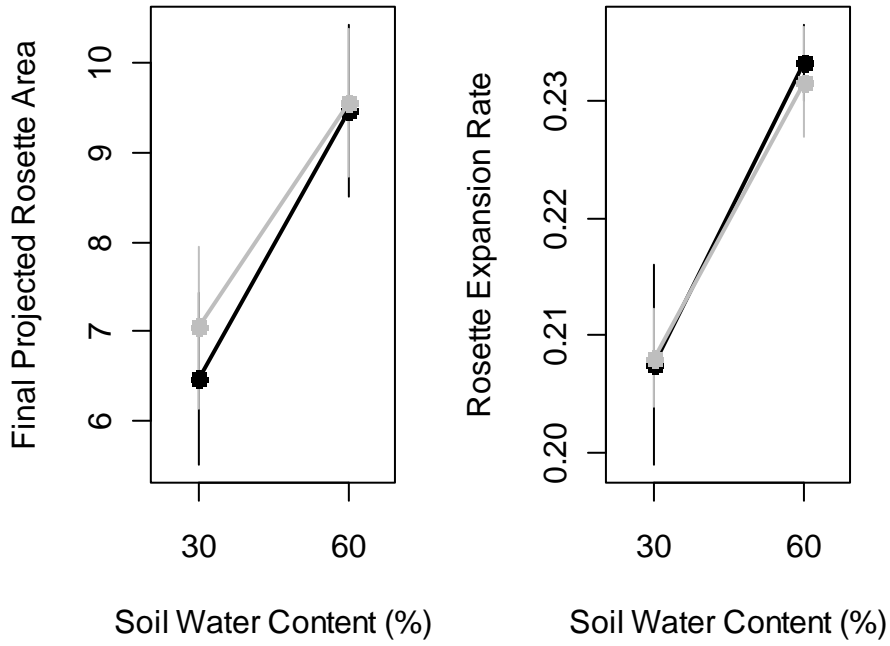
**8HV344**



# Supplementary Figure S10

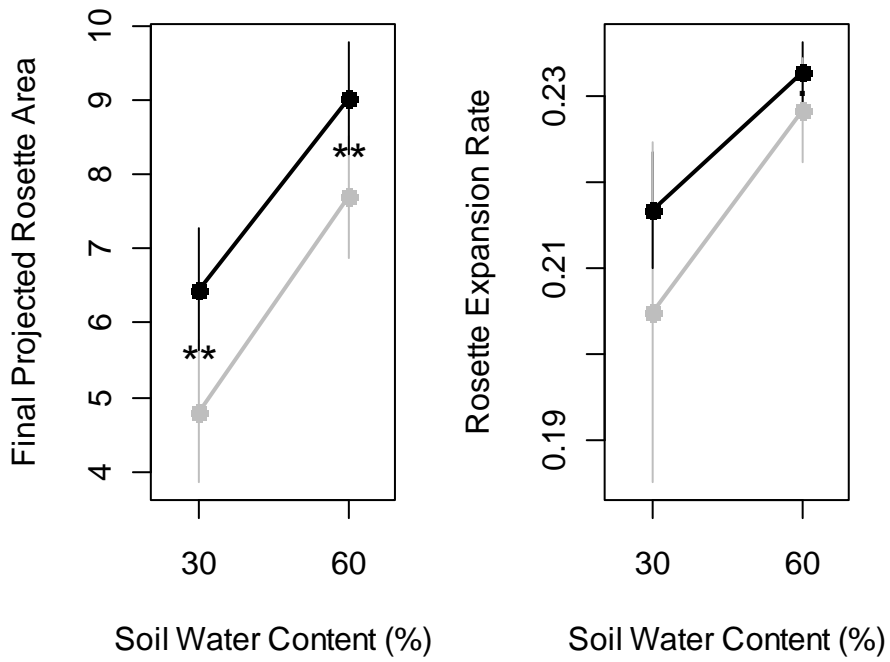
(e)

**8HV218**



(f)

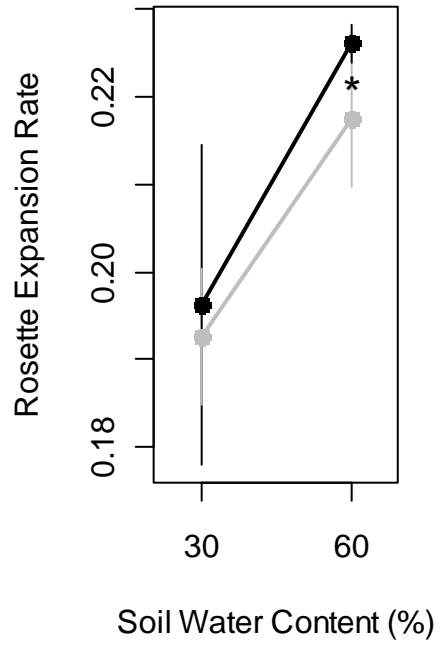
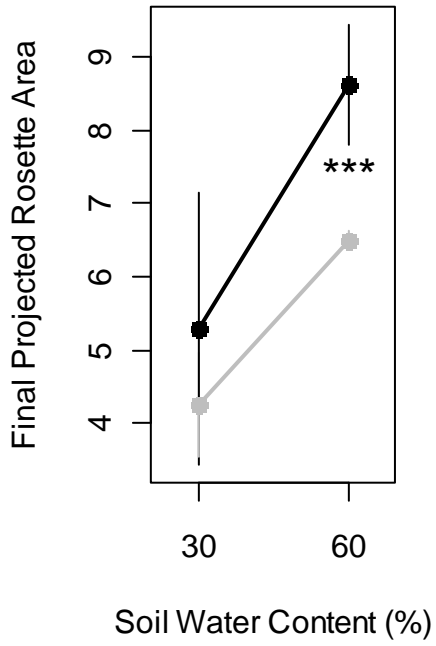
**8HV053**



# Supplementary Figure S10

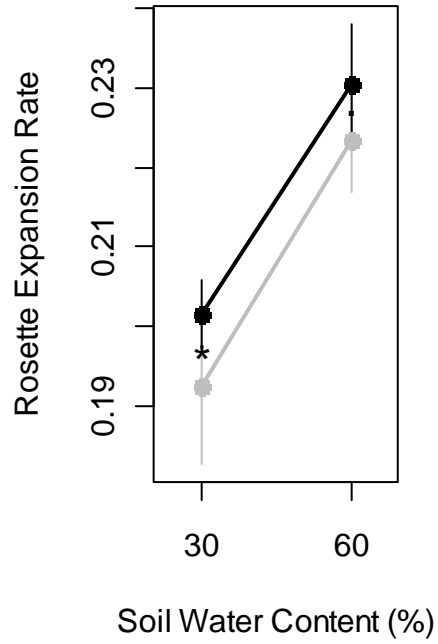
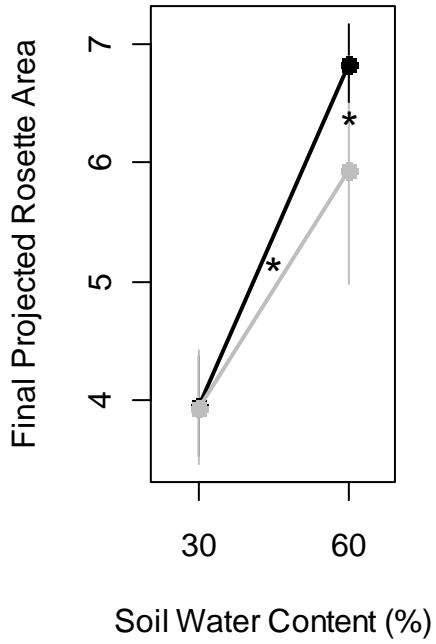
(g)

**8HV169**



(h)

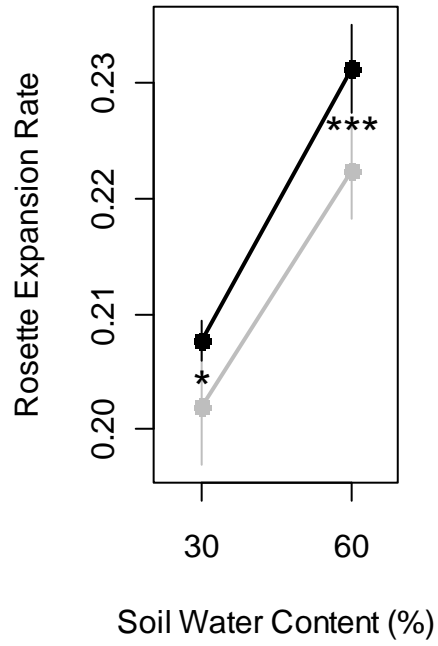
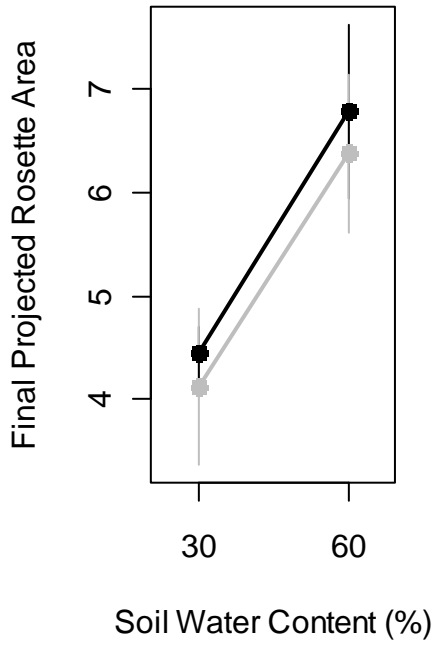
**8HV109**



# Supplementary Figure S10

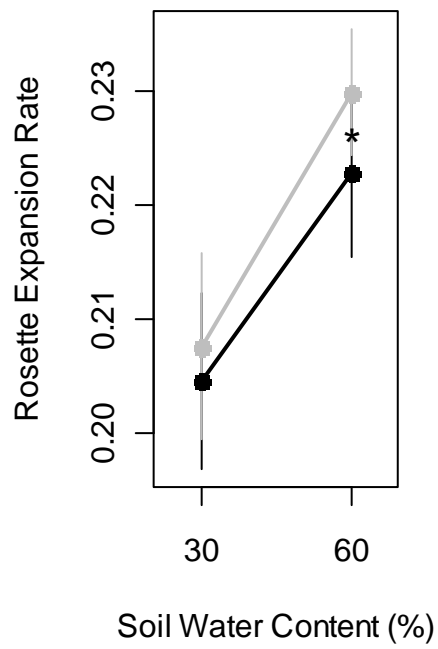
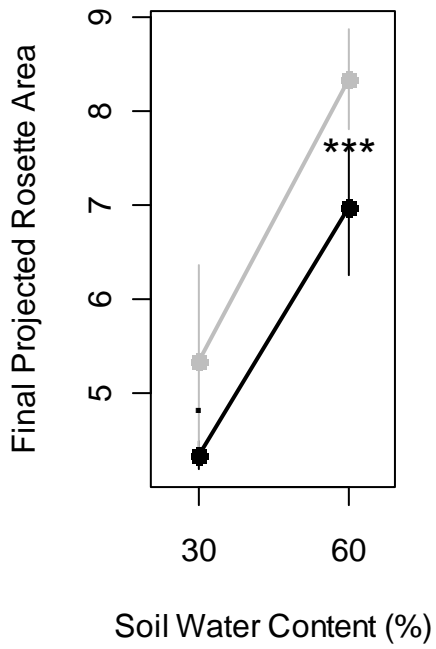
(i)

**8HV236**



(j)

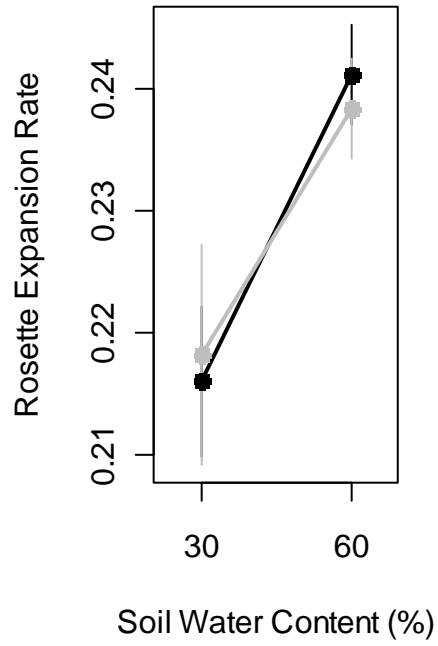
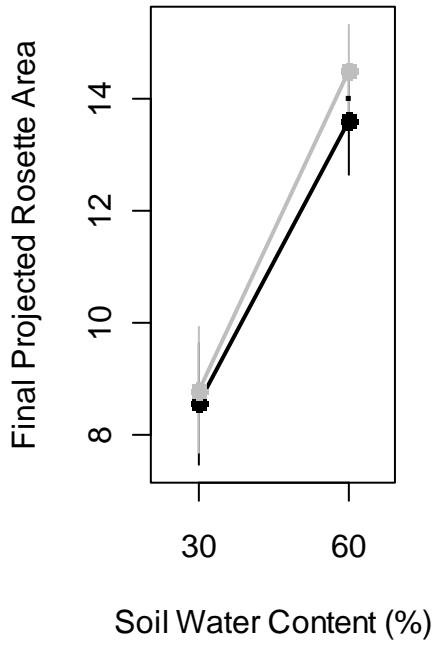
**8HV182**



Supplementary Figure S10

(k)

**8HV291**



(l)

**8HV047**

