Data Article

# An image dataset of diverse safflower (*Carthamus tinctorius* L.) genotypes for salt response phenotyping

Emily Thoday-Kennedy [a], Adam M. Dimech [b], Sameer Joshi [a], Hans D. Daetwyler [b], David Hudson [c], German Spangenberg [b,d], Matthew Hayden [b,d], Surya Kant [a,b,d,*]

[a] *Agriculture Victoria, Grains Innovation Park, Horsham, Victoria, Australia*
[b] *Agriculture Victoria, AgriBio, Centre for AgriBioscience, Bundoora, Victoria, Australia*
[c] *GO Resources Pty Ltd, 15 Sutherland Street, Brunswick, VIC, Australia*
[d] *School of Applied Systems Biology, La Trobe University, Bundoora, Victoria, Australia*

## ARTICLE INFO

## ABSTRACT

This article describes a dataset of high-resolution visible-spectrum images of safflower (*Carthamus tinctorius* L.) plants obtained from a LemnaTec Scanalyser automated phenomics platform along with the associated image analysis output and manually acquired biomass data. This series contains 1832 images of 200 diverse safflower genotypes, acquired at the Plant Phenomics Victoria, Horsham, Victoria, Australia. Two Prosilica GT RGB (red-green-blue) cameras were used to generate 6576 × 4384 pixel portable network graphic (PNG) images. Safflower genotypes were either subjected to a salt treatment (250 mM NaCl) or grown as a control (0 mM NaCl) and imaged daily from 15 to 36 days after sowing. Each snapshot consists of four images collected at a point in time; one of which is taken from above (top-view) and the remainder from the side at either 0°, 120° or 240°. The dataset also includes analysis output quantifying traits and describing phenotypes, as well as manually collected biomass and leaf ion content data. The usage of the dataset is already demonstrated in Thoday-Kennedy *et al.* (2021) [1]. This dataset describes the early growth differences of diverse saf-

* Corresponding author.
    *E-mail address:* surya.kant@agriculture.vic.gov.au (S. Kant).

flower genotypes and identified genotypes tolerant or susceptible to salinity stress. This dataset provides detailed image analysis parameters for phenotyping a large population of safflower that can be used for the training of image-based trait identification pipelines for a wide range of crop species.

Crown Copyright © 2022 Published by Elsevier Inc.

## Specifications Table

| | |
|---|---|
| Subject | Agronomy and Crop Science |
| Specific subject area | Image dataset and analysis output for safflower salinity tolerance |
| Type of data | RGB images (24-bit, PNG format) |
| | Tables with analysed parameters and manual measurements |
| How data were acquired | Side and top mounted 28.8 Megapixel RGB cameras, model Prosilica GT 6600C (Allied Vision Technologies, Stadtroda, Germany) |
| | Safflower RGB image analysis pipeline developed in LemnaGrid software (LemnaTec GmbH, Aachen, Germany) |
| | Flame photometer (Sherwood 420, Sherwood Scientific, Cambridge, United Kingdom) |
| Data format | Portable Network Graphic, 24-bit RGB |
| | Analysed |
| Parameters for data collection | Plants grown in glasshouse with natural light conditions, maintaining 24/15°C day/night temperatures. |
| | Images captured at 0°, 120° & 240° side angle and top view. |
| | Biomass and leaf samples dried for 3 days at 70°C. |
| | Dried leaf samples digested in 10 mL of 1% (v/v) nitric acid for 4 hr in a water bath (TWB-48D, Thermoline Scientific Equipment Pty. Ltd., NSW, Australia). |
| | Ion concentrations determined using standards, 25 µM $Na^+$/ 1500 µM $K^+$ (second leaf pair) at 1/40 or 1/80 dilutions, or 100 µM $Na^+$/ 1000 µM $K^+$ (youngest leaf pair) at 1/10 or 1/20 dilutions. |
| Description of data collection | The dataset is composed of 1832 high-resolution (6576 × 4384 pixels) RGB images and associated data, captured using two RGB cameras in the LemnaTec Scanalyzer 3D automated phenomics platform at Plant Phenomics Victoria, Horsham, Australia. Images represent growth and associated parameters at 36 days after sowing (DAS) with half the plants grown under control conditions (0 mM NaCl) and the other half under salt treatment (250 mM NaCl). Image analysis output based on a custom LemnaGrid safflower image analysis pipeline was obtained for each plant. |
| | Fresh and dry biomass and ion content was manually collected for each plant. |
| Data source location | Institution: Plant Phenomics Victoria, Horsham, Grains Innovation Park, Agriculture Victoria, Department of Jobs, Precincts and Regions. |
| | City/Town/Region: Horsham, Victoria, Australia |
| | Latitude and longitude (and GPS coordinates, if possible) for collected samples/data: 36° 43' 13.67" S, 142° 10' 25.63" E |
| Data accessibility | Repository name: Harvard Dataverse |
| | Direct URL to data: https://dataverse.harvard.edu/dataverse/H2018006 |
| Related research article | E. Thoday-Kennedy, S. Joshi, H.D. Daetwyler, M. Hayden, D. Hudson, G. Spangenberg, S. Kant, Digital phenotyping to delineate salinity response in safflower genotypes, *Frontiers in Plant Science* (2021) *12*, 1196 https://doi.org/10.3389/fpls.2021.662498 |

**Value of the Data**

- This dataset is a collection of RGB images and associated image analysis parameters describing multiple traits contributing to diverse phenotypes as well as manually collected biomass and ion content data. The data set can be used to understand differences in early vegetative growth for diverse safflower genotypes and quantify variation in their response to salinity.
- This dataset can be a valuable resource to be utilised by researchers seeking to elucidate early vegetative physiological responses of a diverse collection of safflower genotypes and their responses to salinity.
- These data can be used/reused to compare results on the growth and physiology of diverse safflower genotypes under varying abiotic or biotic stresses, as well as used as phenotypic data in both forward and reverse genomic studies. This dataset can also be used in the development and refinement of high-throughput image-analysis pipelines for a wide range of crop species.
- This dataset of high-quality RGB images and associated physiological parameters is a unique resource for the further understanding of an underutilised but valuable safflower crop [2].

## 1. Objective

Safflower has previously been reported as a salt tolerant crop, although responses vary across genotypes [3]. To better understand the existing diversity of responses in the Agriculture Victoria safflower collection, a high-throughput digital phenotyping experiment was designed to screen the collection for salt tolerance. The dataset described in this manuscript was generated during this process to screen diverse genotypes for physiological responses to salt treatment during early vegetative growth. This dataset supports our previous publication [1], by expanding on the individual genotypic responses to control and salt treatments, as well as the physiological, phenological and biochemical parameters captured for each genotype.

## 2. Data Description

This dataset consists of 1832 RGB images in PNG format along with associated metadata and results data files in three directories accordingly labelled.

Images, in the 'images' folder, and metadata, in the 'metadata' folder, are organised by barcode identifiers (i.e. H20180061010MM) consisting of internal experimental reference "H2018006", a genotype ID corresponding to data in the results files, i.e. "101", salt treatment classification "0MM or 250MM", then a date and time stamp, internal experiment name and the angle image was captured at (VIS-0 = 0° side view, VIS-120 = 120° side view, VIS-240 = 240° side view, VIS-TV = top view). Fig. 1 shows an example of four typical images from each snapshot captured for a single plant grown under either treatment.

The 'results' folder contains two files which describe either image analysis output (CSV file) or manually collected data (Excel sheet) for 458 plants or 200 genotypes grown under both control and salt treatments. Table 1 describes the output parameters derived from image analysis for each of the four images per plant.
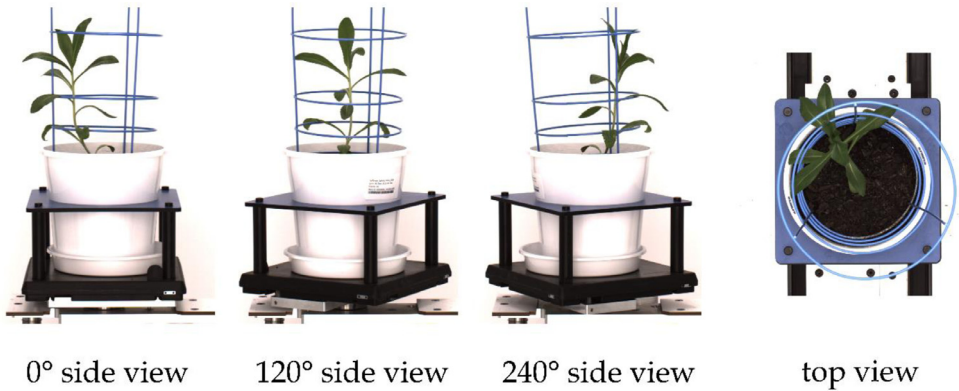
**Fig. 1.** Example of parts of high-resolution RGB images captured at four separate angles (three side views and one top view) for a single plant at 36 days after sowing.

**Table 1**
Definitions of image analysis output parameters.

| Parameter | Description |
| --- | --- |
| Snapshot ID | Unique barcode for each plant in experiment |
| Genotype | Corresponding genotype ID |
| Treatment | Salinity treatment class: control (0 mM Na$^+$) or salt (250 mM Na$^+$) |
| 2nd Moment Principle Axis Large Abs | Sum of distance in pixels of larger axis from the centre mass |
| 2nd Moment Principle Axis Small Abs | Sum of distance in pixels of smaller axis from the centre mass |
| Area | Area of region-of-interest, expressed in pixels |
| Boundary Point Count | Perimeter of region-of-interest including holes |
| Boundary Point Roundness | Ratio of boundary count$^2$: area |
| Boundary Points To Area Ratio | Surface coverage ratio of boundary count to area |
| Caliper Length | Maximum length of the identified plant expressed as pixels |
| Circumference | Perimeter of region-of-interest excluding holes |
| Compactness | Ratio of area to convex hull area |
| Convex Hull Area | Area of the smallest perimeter with no concave sections that contains the region-of-interest, expressed in pixels |
| Convex Hull Circumference | Length of the convex hull |
| Excentricity | Extent of radial symmetry of the region-of-interest spread, generally the value is between 0 and 1 |
| Min Enclosing Circle Diameter | Diameter of the smallest circle enclosing the region-of-interest |
| Min Area Rectangle Area | Minimum rectangular area that encloses the region-of-interest, expressed in pixels |
| Result Overviewgreen.ColorClassAreaAbsolut | Area of the region-of-interest where pixels were binned as "green", expressed in pixels |
| Result Overviewgreen.ColorClassAreaRelative | Area of the region-of-interest where pixels were binned as "green", expressed as a ratio to total area |
| Result Overviewyellow/brown.ColorClassAreaAbsolut | Area of the region-of-interest where pixels were binned as "yellow/brown", expressed in pixels |
| Result Overviewyellow/brown.ColorClassAreaRelative | Area of the region-of-interest where pixels were binned as "yellow/brown", expressed as a ratio to area |
| Roundness | Ratio of circumference$^2$ to area |

## 3. Experimental Design, Materials and Methods

### 3.1. Plant Growth

Images and associated data were acquired at the Plant Phenomics Victoria, Horsham (PPVH) an automated, high-throughput plant imaging facility operated by Agriculture Victoria, Department of Jobs, Precincts and Regions. This facility contains a LemnaTec Scanalyzer 3D plant-to-

sensor platform (Lemnatec GmbH, Aachen, Germany), consisting of a 600 carrier capacity conveyor system, with automated weighing and watering stations, and a digital imaging cabinet containing two Prosilica GT visible-spectrum RGB cameras. Two hundred diverse safflower genotypes, representing the maximum genetic diversity in the Agriculture Victoria safflower collection (Supp. Table 1; [1]), were phenotyped in the facility to further understand their salinity tolerance. Safflower seeds were directly sown in 200 mm diameter white plastic pots containing pine-based potting mix and additional fertilisers [1]. White pots were used to aid in thresholding steps during image analysis. Blue powder-coated steel metal cages were used to support plants as they grew. Plants were grown in controlled glasshouse conditions at 24/15°C (day/night) under natural light.

At 17 days after sowing (DAS) when the first leaf pair was unfolding (growth stage BBCH GS12; [4]) a treatment consisting either of a 0 mM NaCl (control) or 250 mM NaCl solution was applied. To prevent osmotic shock, 150 mL of each treatment solution was applied via the saucer over 2 days for the 0 mM treatment and 3 days for the 250 mM treatment. To ensure calcium activity remained the same between treatments, 33 mM of $CaCl_2$ was added to the master 1M NaCl stock solution, which was diluted to achieve the correct application, based on the gravimetric soil water content. Moisture content in the pots was maintained at 80% field capacity throughout the experiment.

### 3.2. Image Capture and Analysis

Each plant was loaded into a carrier containing a radiofrequency identification chip for tracking. During imaging plants were moved into the RGB imaging cabinet, in which the side view camera was positioned 1.1 m from the pot and the top view camera was 3.6 m from the pot. Illumination was provided by multiple 6500K T5 fluorescent lights positioned on the sides and top of the imaging cabinet to provide a clear imaging environment with minimal shadows. Images were then captured using two fixed-zoom RGB cameras (top and side mounted), Prosilica GT 6600C (Allied Vision Technologies, Stadtroda, Germany). Camera specifications are 28.8 Megapixel, 6576 × 4384 pixels, 14-bit resolution, F-Mount, 35 mm CCD progressive monochrome and colour, frame rate of 4 fps, and cell size of 5.5 um. Using the camera mounted directly above the plants, one digital RGB top-view image was acquired. The three side view images were captured after consecutive rotations of the 'lifter and turner' at 0°, 120° and 240°, with plants elevated to 30 cm. Captured images were stored as Bayer GR8-format image blobs in a PostgreSQL database.

Images were analysed using an image analysis pipeline developed in the LemnaGrid software (Lemnatec GmbH, Aachen, Germany). Demosaicing was initially carried out on the image blobs to reconstruct the full-colour images in PNG format. Colour images were split into HSV and L*a*b colour spaces [5] and a series of thresholding steps applied to isolate the foreground (region-of-interest) from background. Further adaptive thresholding process were applied, as well as a median filter, edge-smoothing and opening steps to remove unwanted solitary pixels (noise). Finally, image object composition combined any spatially independent objects to create a single identified object, the plant. Colour binning using the nearest-neighbour method was applied to delineate green and non-green ("yellow/brown") pixels. Morphological measurements were extracted automatically from the resultant region-of-interest.

### 3.3. Manual Destructive Harvest and Ion Analysis

All plants were destructively harvested 36 DAS to determine fresh and dry shoot biomass. Depending on treatment and genotype, plants were between growth stages BBCH GS 16 and GS 37 [4]. Plants were cut at soil level and weighed. The third and fourth leaves (second true leaf pair) and first and second-youngest fully-expanded leaves (youngest leaf pair) were removed,

weighed separately and the pairs were put into separate 15 mL tubes. All fresh biomass and two leaf pair samples were dried for 3 days at 70°C, then weighed to obtain dry mass.

To obtain sodium ($Na^+$) and potassium ($K^+$) concentrations of the leaf pairs, leaves were digested in 10 mL of 1% (v/v) nitric acid solution in a water bath (TWB-48D, Thermoline Scientific Equipment Pty. Ltd., NSW, Australia) at 100°C for 4 h in a water bath. Digests were used in flame photometry, using a Sherwood 420 Flame Photometer (Sherwood Scientific, Cambridge, UK) to determine ion content. Based on original leaf pair weights, samples of the digests were diluted in distilled water to a volume of 2 mL, and then passed through the flame photometer against known standards. For the second leaf pair a standard of 25 μM $Na^+$/ 1500 μM $K^+$ was used, with most samples diluted at a 1/40 or 1/80 ratio, while for the youngest leaf pair a standard of 100 μM $Na^+$/ 1000 μM $K^+$ was used at 1/10 or 1/20 dilutions.

## Ethics Statement

This study does not involve experiments on humans or animals.

## CRediT Author Statement

**Emily Thoday-Kennedy:** Data curation, Methodology, Investigation, Writing – original draft preparation; **Adam Dimech:** Data curation, Writing – review & editing; **Sameer Joshi:** Investigation, Data curation; **David Hudson:** Conceptualisation, Reviewing and editing; **Hans D Daetwyler:** Conceptualisation, Reviewing and editing; **German Spangenberg:** Conceptualisation, Funding, Reviewing and editing; **Matthew Hayden:** Conceptualisation, Funding, Writing – review & editing; **Surya Kant:** Conceptualisation, Methodology, Supervision, Writing – review & editing.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have or could be perceived to have influenced the work reported in this article.

## Data Availability

Digital Phenotyping to Delineate Salinity Response in Safflower Genotypes (Original data) (Dataverse).

## Acknowledgments

## References

[1] E. Thoday-Kennedy, S. Joshi, H.D. Daetwyler, M. Hayden, D. Hudson, G. Spangenberg, S. Kant, Digital phenotyping to delineate salinity response in safflower genotypes, Front. Plant Sci. 12 (2021) 662498, doi:10.3389/fpls.2021.662498.
[2] L. Dajue, H.-H. Mündel, *Safflower. Carthamus tinctorius L.* Promoting the conservation and use of underutilized and neglected crops, Institute of Plant Genetics and Crop Plant Research, Gatersleben/International Plant Genetic Resources Institute, Rome, Italy, 1996 Vol. 7.

[3] M. Ashraf, H. Fatima, Responses of salt-tolerant and salt-sensitive lines of safflower (Carthamus tinctorius L.) to salt stress, Acta Physiol. Plantarum 17 (1) (1995) 61–70.
[4] A.C. Flemmer, M.C. Franchini, L.I. Lindström, Description of safflower (Carthamus tinctorius) phenological growth stages according to the extended BBCH scale, Ann. Appl. Biol. 166 (2) (2015) 331–339, doi:10.1111/aab.12186.
[5] M.R. Luo, CIELAB, in: R. Luo (Ed.), Encyclopedia of Color Science and Technology, Springer Berlin Heidelberg, Berlin, Heidelberg, 2014, pp. 1–7.