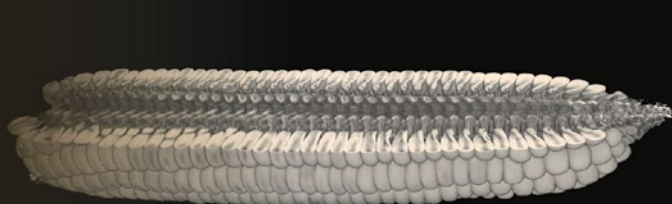


Methods in
Molecular Biology 2539

Springer Protocols



Argelia Lorence
Karina Medina Jimenez *Editors*

High-Throughput Plant Phenotyping

Methods and Protocols

 Humana Press

METHODS IN MOLECULAR BIOLOGY

Series Editor

John M. Walker

School of Life and Medical Sciences

University of Hertfordshire

Hatfield, Hertfordshire, UK

For further volumes:

<http://www.springer.com/series/7651>

For over 35 years, biological scientists have come to rely on the research protocols and methodologies in the critically acclaimed *Methods in Molecular Biology* series. The series was the first to introduce the step-by-step protocols approach that has become the standard in all biomedical protocol publishing. Each protocol is provided in readily-reproducible step-by-step fashion, opening with an introductory overview, a list of the materials and reagents needed to complete the experiment, and followed by a detailed procedure that is supported with a helpful notes section offering tips and tricks of the trade as well as troubleshooting advice. These hallmark features were introduced by series editor Dr. John Walker and constitute the key ingredient in each and every volume of the *Methods in Molecular Biology* series. Tested and trusted, comprehensive and reliable, all protocols from the series are indexed in PubMed.

High-Throughput Plant Phenotyping

Methods and Protocols

Edited by

Argelia Lorence

Arkansas Biosciences Institute, Arkansas State University, Jonesboro, AR, USA

Karina Medina-Jimenez

Arkansas Biosciences Institute, Arkansas State University, Jonesboro, AR, USA

Editors

Argelia Lorence
Arkansas Biosciences Institute
Arkansas State University
Jonesboro, AR, USA

Karina Medina-Jimenez
Arkansas Biosciences Institute
Arkansas State University
Jonesboro, AR, USA

ISSN 1064-3745

Methods in Molecular Biology

ISBN 978-1-0716-2536-1

<https://doi.org/10.1007/978-1-0716-2537-8>

ISSN 1940-6029 (electronic)

ISBN 978-1-0716-2537-8 (eBook)

© Springer Science+Business Media, LLC, part of Springer Nature 2022

Chapter 21 is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>). For further details see license information in the chapter.

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Humana imprint is published by the registered company Springer Science+Business Media, LLC, part of Springer Nature.

The registered company address is: 1 New York Plaza, New York, NY 10004, U.S.A.

Preface

Increasing the productivity of crops is imperative to satisfy the growing demand for food, feed, and fuel in the world. The development of crops with higher yields capable of thriving under adverse environmental conditions requires novel strategies that enable better and faster assessment of the genome-by-environment-by-management (GxExM) interactions. Plant phenomics or high throughput plant phenotyping (HTPP) consists in the application and development of different methodologies to capture information related with performance, function, and structure of a large number of plants. The main purpose of plant phenomics is to understand plant behavior under a vast variety of scenarios and how exactly the genotypic traits are expressed through the plant phenotype.

In *High-Throughput Plant Phenotyping: Review and Protocols*, readers will find a collection of state-of-the-art, step-by-step, and reproducible protocols to quantify the GxExM interactions in a variety of model and plant crops. Reflecting the multidisciplinary nature of this area of research, the book contains all aspects that are key to HTPP experiments including plant growth and care, experimental design considerations, image acquisition tools and robots, novel algorithms for image analysis, and protocols discussing statistical and network analysis. The book is divided into five parts. In the first one, leaders in the field contributed HTPP protocols for plants growing under controlled conditions. In the second part, we present novel algorithms for extracting data from seed images, color analysis from fruits, and other digital readouts from 2D objects. Part III is comprised of two chapters describing keys to the success of molecular imaging protocols using PET and X-ray approaches. The fourth part is a collection of HTPP protocols for crops growing under field conditions. Part V contains chapters dedicated to molecular analysis, metabolomics, network analysis, and statistical methods for the quantitative genetic analysis of HTP data.

We are grateful to all the talented people that contributed their expertise and know-how in these excellent chapters. These colleagues, in addition to top-quality scientists, are also compassionate people. I would like to thank all contributors for the patience and grace you offered me during the difficult months when my mother was very ill and the ones after her passing.

Mom, you are the strongest and wisest woman I know. I aspire to follow your example and make you proud. This book is for you.

Jonesboro, AR, USA

Argelia Lorence

Contents

<i>Preface</i>	<i>v</i>
<i>Contributors</i>	<i>ix</i>

PART I HTP PROTOCOLS FOR PLANTS GROWING UNDER CONTROLLED CONDITIONS

1 High-Throughput Screening to Examine the Dynamic of Stay-Green by an Imaging System	3
<i>Daniel Padilla-Chacón and Cecilia B. Peña-Valdivia</i>	
2 An Automated High-Throughput Phenotyping System for <i>Marchantia polymorpha</i>	11
<i>Karina Medina-Jimenez, Mario A. Arteaga-Vazquez, and Argelia Lorence</i>	
3 A Novel High-Throughput Phenotyping Hydroponic System for Nitrogen Deficiency Studies in <i>Arabidopsis thaliana</i>	19
<i>Lucia M. Acosta-Gamboa, Zachary C. Campbell, Fei Gao, Benjamin Babst, and Argelia Lorence</i>	
4 <i>Camelina sativa</i> High-Throughput Phenotyping Under Normal and Salt Conditions Using a Plant Phenomics Platform	25
<i>Emilio Vello, John Aguirre, Yang Shao, and Thomas Bureau</i>	
5 A Straightforward High-Throughput Aboveground Phenotyping Platform for Small- to Medium-Sized Plants	37
<i>Denise Caldwell and Anjali S. Iyer-Pascuzzi</i>	
6 Wireless Fixed Camera Network for Greenhouse-Based Plant Phenotyping	49
<i>Nadia Shakoor and Todd C. Mockler</i>	
7 Experimental Design for Controlled Environment High-Throughput Plant Phenotyping	57
<i>Jennifer L. Clarke, Yumou Qin, and James C. Schnable</i>	

PART II NOVEL ALGORITHMS FOR HTP

8 High-Throughput Extraction of Seed Traits Using Image Acquisition and Analysis	71
<i>Chongyuan Zhang and Sindhuja Sankaran</i>	
9 ColourQuant: A High-Throughput Technique to Extract and Quantify Color Phenotypes from Plant Images	77
<i>Mao Li, Margaret H. Frank, and Zoë Migicovsky</i>	

10	Using Cameras for Precise Measurement of Two-Dimensional Plant Features: CASS	87
	<i>Amy Tabb, Germán A. Holguín, and Rachel Naegele</i>	
PART III MOLECULAR PLANT IMAGING		
11	Positron Emission Tomography (PET) for Molecular Plant Imaging	97
	<i>Sergey Komarov and Yuan-Chuan Tai</i>	
12	Phenotyping Complex Plant Structures with a Large Format Industrial Scale High-Resolution X-Ray Tomography Instrument	119
	<i>Keith E. Duncan and Christopher N. Topp</i>	
PART IV HTP PROTOCOLS FOR PLANTS GROWING UNDER FIELD CONDITIONS		
13	Challenges for a Massive Implementation of Phenomics in Plant Breeding Programs	135
	<i>Gustavo A. Lobos, Félix Estrada, Alejandro del Pozo, Sebastián Romero-Bravo, Cesar A. Astudillo, and Freddy Mora-Poblete</i>	
14	Designing Experiments for Physiological Phenomics	159
	<i>Addie Thompson, Michael Kantar, and Katy Rainey</i>	
15	Design Considerations for In-Field Measurement of Plant Architecture Traits Using Ground-Based Platforms	171
	<i>Piyush Pandey and Sierra Young</i>	
16	Design and Construction of Unmanned Ground Vehicles for Sub-canopy Plant Phenotyping	191
	<i>Adam Stager, Herbert G. Tanner, and Erin Sparks</i>	
17	Nighttime Chlorophyll Fluorescence Imaging of Dark-Adapted Plants Using a Robotic Field Phenotyping Platform	213
	<i>Maria Newcomb and Nadia Shakoov</i>	
PART V MOLECULAR, METABOLOMICS, NETWORK ANALYSIS, AND QUANTITATIVE GENETIC ANALYSIS OF HTP DATA		
18	A Method for Rapid and Reliable Molecular Detection of Drought-Response Genes in <i>Sorghum bicolor</i> (L.) Moench Roots	223
	<i>Juan B. Fontanet-Manzanique, David Blasco-Escámez, Damiano Martignago, Andrés Rico-Medina, and Ana I. Caño-Delgado</i>	
19	High-Throughput Profiling of Metabolic Phenotypes Using High-Resolution GC-MS	235
	<i>Nishikant Wase, Nathan Abshire, and Toshihiro Obata</i>	
20	Gene Co-expression Network Analysis and Linking Modules to Phenotyping Response in Plants	261
	<i>Qian Du, Malachy T. Campbell, Huibei Yu, Kan Liu, Harkamal Walia, Qi Zhang, and Chi Zhang</i>	

21	Statistical Methods for the Quantitative Genetic Analysis of High-Throughput Phenotyping Data	269
	<i>Gota Morota, Diego Jarquin, Malachy T. Campbell, and Hiroyoshi Iwata</i>	
	<i>Index</i>	297

Contributors

- NATHAN ABSHIRE • *Department of Biochemistry and Center for Plant Science Innovation, University of Nebraska-Lincoln, Lincoln, NE, USA*
- LUCIA M. ACOSTA-GAMBOA • *Arkansas Biosciences Institute, Arkansas State University, Jonesboro, AR, USA*
- JOHN AGUIRRE • *Department of Biology, McGill University, Montreal, QC, Canada*
- MARIO A. ARTEAGA-VAZQUEZ • *Instituto de Biotecnología y Ecología Aplicada (INBIOTECA), Universidad Veracruzana, Xalapa, Mexico*
- CESAR A. ASTUDILLO • *Department of Computer Science, Faculty of Engineering, Universidad de Talca, Curico, Chile*
- BENJAMIN BABST • *Arkansas Forest Resources Center, Division of Agriculture, University of Arkansas System, Monticello, AR, USA*
- DAVID BLASCO-ESCÁMEZ • *Department of Molecular Genetics, Centre for Research in Agricultural Genomics CSIC-IRTA-UAB-UB (CRAG), Barcelona, Spain*
- THOMAS BUREAU • *Department of Biology, McGill University, Montreal, QC, Canada*
- DENISE CALDWELL • *Department of Botany and Plant Pathology and Center for Plant Biology, Purdue University, West Lafayette, IN, USA*
- MALACHY T. CAMPBELL • *Department of Agronomy and Horticulture, Center for Plant Science and Innovation, University of Nebraska, Lincoln, NE, USA; Department of Animal and Poultry Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA*
- ZACHARY C. CAMPBELL • *Arkansas Biosciences Institute, Arkansas State University, Jonesboro, AR, USA*
- ANA I. CAÑO-DELGADO • *Department of Molecular Genetics, Centre for Research in Agricultural Genomics CSIC-IRTA-UAB-UB (CRAG), Barcelona, Spain*
- JENNIFER L. CLARKE • *Department of Statistics, University of Nebraska, Lincoln, NE, USA*
- ALEJANDRO DEL POZO • *Plant Breeding and Phenomics Center, Faculty of Agricultural Sciences, Universidad de Talca, Talca, Chile*
- QIAN DU • *School of Biological Sciences, Center for Plant Science and Innovation, University of Nebraska, Lincoln, NE, USA*
- KEITH E. DUNCAN • *Donald Danforth Plant Science Center, Saint Louis, MO, USA*
- FÉLIX ESTRADA • *Plant Breeding and Phenomics Center, Faculty of Agricultural Sciences, Universidad de Talca, Talca, Chile*
- JUAN B. FONTANET-MANZANEQUE • *Department of Molecular Genetics, Centre for Research in Agricultural Genomics CSIC-IRTA-UAB-UB (CRAG), Barcelona, Spain*
- MARGARET H. FRANK • *Plant Biology Section, School of Integrative Plant Science, Cornell University, Ithaca, NY, USA*
- FEI GAO • *Arkansas Forest Resources Center, Division of Agriculture, University of Arkansas System, Monticello, AR, USA*
- GERMÁN A. HOLGUÍN • *Electrical Engineering Department, Universidad Tecnológica de Pereira, Pereira, Colombia*
- HIROYOSHI IWATA • *Department of Agricultural and Environmental Biology, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, Japan*

- ANJALI S. IYER-PASCUZZI • *Department of Botany and Plant Pathology and Center for Plant Biology, Purdue University, West Lafayette, IN, USA*
- DIEGO JARQUIN • *Agronomy Department, University of Florida, Gainesville, FL, USA*
- MICHAEL KANTAR • *Department of Tropical Plant and Soil Science, University of Hawaii at Manoa, Honolulu, HI, USA*
- SERGEY KOMAROV • *Department of Radiology, Washington University in St. Louis, St. Louis, MO, USA*
- MAO LI • *Donald Danforth Plant Science Center, St. Louis, MO, USA*
- KAN LIU • *School of Biological Sciences, Center for Plant Science and Innovation, University of Nebraska, Lincoln, NE, USA*
- GUSTAVO A. LOBOS • *Plant Breeding and Phenomics Center, Faculty of Agricultural Sciences, Universidad de Talca, Talca, Chile*
- ARGELIA LORENCE • *Arkansas Biosciences Institute, Arkansas State University, Jonesboro, AR, USA; Department of Chemistry and Physics, Arkansas State University, Jonesboro, AR, USA*
- DAMIANO MARTIGNAGO • *Department of Molecular Genetics, Centre for Research in Agricultural Genomics CSIC-IRTA-UAB-UB (CRAG), Barcelona, Spain; Department of Biosciences, University of Milan, Milan, Italy*
- KARINA MEDINA-JIMENEZ • *Arkansas Biosciences Institute, Arkansas State University, Jonesboro, AR, USA*
- ZOE MIGICOVSKY • *Department of Plant, Food, and Environmental Sciences, Faculty of Agriculture, Dalhousie University, Truro, NS, Canada*
- TODD C. MOCKLER • *Donald Danforth Plant Science Center, St. Louis, MO, USA*
- FREDDY MORA-POBLETE • *Institute of Biological Sciences, Universidad de Talca, Talca, Chile*
- GOTA MOROTA • *Department of Animal and Poultry Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA*
- RACHEL NAEGELE • *United States Department of Agriculture, Agricultural Research Service, Sugarbeet and Bean Research Unit (USDA-ARS), East Lansing, MI, USA*
- MARIA NEWCOMB • *University of Arizona, School of Plant Sciences, Maricopa Agricultural Center, Maricopa, AZ, USA*
- TOSHIHIRO OBATA • *Department of Biochemistry and Center for Plant Science Innovation, University of Nebraska-Lincoln, Lincoln, NE, USA*
- DANIEL PADILLA-CHACÓN • *Programa de Posgrado en Botánica, CONACyT-Colegio de Postgraduados. Km 36.5 Carretera Mexico-Texcoco, Montecillo, MX, Mexico*
- PIYUSH PANDEY • *Department of Biological and Agricultural Engineering, North Carolina State University, Raleigh, NC, USA*
- CECILIA B. PEÑA-VALDIVIA • *Programa de Posgrado en Botánica, Colegio de Postgraduados. Km 36.5 Carretera México-Texcoco, Montecillo, MX, Mexico*
- YUMOU QIU • *Department of Statistics, Iowa State University, Ames, IA, USA*
- KATY RAINEY • *Department of Agronomy, Purdue University, West Lafayette, IN, USA*
- ANDRÉS RICO-MEDINA • *Department of Molecular Genetics, Centre for Research in Agricultural Genomics CSIC-IRTA-UAB-UB (CRAG), Barcelona, Spain*
- SEBASTIÁN ROMERO-BRAVO • *Department of Agricultural Sciences, Universidad Católica del Maule, Curico, Chile*
- SINDHUJA SANKARAN • *Department of Biological Systems Engineering, Washington State University, Pullman, WA, USA*
- JAMES C. SCHNABLE • *Department of Agronomy and Horticulture, Center for Plant Science Innovation, University of Nebraska, Lincoln, NE, USA*

- NADIA SHAKOOR • *Donald Danforth Plant Science Center, St. Louis, MO, USA*
- YANG SHAO • *Department of Biology, McGill University, Montreal, QC, Canada*
- ERIN SPARKS • *Department of Plant and Soil Sciences, University of Delaware, Newark, DE, USA*
- ADAM STAGER • *Department of Mechanical Engineering, University of Delaware, Newark, DE, USA; TRIC Robotics LLC, Newark, DE, USA*
- AMY TABB • *United States Department of Agriculture, Agricultural Research Service, Appalachian Fruit Research Station (USDA-ARS-AFRS), Kearneysville, WV, USA*
- YUAN-CHUAN TAI • *Department of Radiology, Washington University in St. Louis, St. Louis, MO, USA*
- HERBERT G. TANNER • *Department of Mechanical Engineering, University of Delaware, Newark, DE, USA*
- ADDIE THOMPSON • *Department of Plant, Soil and Microbial Sciences, Michigan State University, East Lansing, MI, USA*
- CHRISTOPHER N. TOPP • *Donald Danforth Plant Science Center, Saint Louis, MO, USA*
- EMILIO VELLO • *Department of Biology, McGill University, Montreal, QC, Canada*
- HARKAMAL WALIA • *Department of Agronomy and Horticulture, Center for Plant Science and Innovation, University of Nebraska, Lincoln, NE, USA*
- NISHIKANT WASE • *Department of Biochemistry and Center for Plant Science Innovation, University of Nebraska-Lincoln, Lincoln, NE, USA; Biomolecular Analysis Facility, University of Virginia School of Medicine, Charlottesville, VA, USA*
- SIERRA YOUNG • *Department of Civil and Environmental Engineering, Utah State University, Logan, UT, USA*
- HUIHUI YU • *School of Biological Sciences, Center for Plant Science and Innovation, University of Nebraska, Lincoln, NE, USA*
- CHI ZHANG • *School of Biological Sciences, Center for Plant Science and Innovation, University of Nebraska, Lincoln, NE, USA*
- CHONGYUAN ZHANG • *Department of Biological Systems Engineering, Washington State University, Pullman, WA, USA*
- QI ZHANG • *Department of Mathematics and Statistics, College of Engineering and Physical Sciences (CEPS), University of New Hampshire, Durham, NH, USA*

Part I

HTP Protocols for Plants Growing Under Controlled Conditions



Chapter 1

High-Throughput Screening to Examine the Dynamic of Stay-Green by an Imaging System

Daniel Padilla-Chacón and Cecilia B. Peña-Valdivia

Abstract

The development of RGB (red, green, blue) sensors has opened the way for plant phenotyping. This is relevant because plant phenotyping allows us to visualize the product of the interaction between the plant ontogeny, anatomy, physiology, and biochemistry. Better yet, this can be achieved at any stage of plant development, i.e., from seedling to maturity. Here, we describe the use of phenotyping, based on the stay-green trait, of common bean (*Phaseolus vulgaris* L.) plant, as a model, stressed by water deficit, to elucidate the result of that interaction. Description is based on interpretation of RGB digital images acquired using a phenomic platform and a specific software. These images allow us to obtain a data group related to the color parameters that quantify the changes and alterations in each plant growth and development.

Key words Stay-green, High-throughput screening, RGB, Common bean, Senescence

1 Introduction

By definition, senescence is a complex trait that reflects the final stage of development during which the plant recycles nutriment, induces gene expression, and involves the interactions of many signaling pathways that may have significant impact on increasing future food production [1, 2].

High-throughput screening (HTS) is a recent scientific method in which hundreds of thousands of experimental samples are subjected to simultaneous testing under given conditions [3]. The sample themselves may take the form of molecules, cells, plants, or field crops to generating large datasets to answer complex biological questions [4]. In particular, “stay-green” or greenless is accepted to be one the most vulnerable parts of the first manifestation of leaf stress. That can be a result of alterations in hormone metabolism and signaling, particularly those affecting networks involving cytokinins and ethylene associated with chlorophyll synthesis-degradation. Symptoms of premature senescence such as

leaf greenless are related to alterations in photosystems (PS) and can be followed up with the chlorophyll fluorescence. Because this fluorescence depends directly on tissue photosynthetic activity, and explains the flow of electrons through PSII, this can be an indicative of photosynthetic efficiency and photosynthetic performance [5–8]. Empirical selection for functional stay-green has contributed to increasing crop yields, particularly where it is part of a strategy that also targets other traits such as sink capacity and environmental sensitivity and is associated with appropriate crop management methodology [8, 9].

Here, we describe a procedure for a rapid, inexpensive, and noninvasive assessment of plants during abiotic stress treatment in relation to chlorophyll turnover and stay-green traits. The changes in major photosynthetic parameters during increasing abiotic stress were monitored via RGB (red, green, blue) imaging in leaves and fruits. The method described here allows the acquisition of quantitative numerical traits that are amenable to statistical analysis. Thus, we applied an accurate method analyzing changes in image time-series to investigate the plant growth and phenotypic response to abiotic stress conditions or different levels of fertilizer of common bean (*Phaseolus vulgaris* L.). A phenotyping platform Scanalyzer PL is used to image in RGB and to monitor plant greenness by phenotypic image analysis accurately measured plant biomass.

2 Materials

1. Diverse growth habit (type I) of *P. vulgaris* cultivars can be used. Each must include a minimum of five replications, each represented by a single plant in a pot (*see Note 1*).
2. Each plant should grow in a plastic pot (20 cm × 12 cm × 19 cm), with 4–6 kg of soil or sand.
3. The researcher must make sure to establish at least two plant groups: one will be the control that will grow with zero level of the stress factor and the other will be maintained with the stress factor (other plant groups can be under more stress levels).
4. Plant images are obtained using a Scanalyzer PL semiautomated platform imaging system (LemnaTec GmbH, Aachen, Germany) (Fig. 1a).

3 Methods

3.1 Image Acquisition (Scanalyzer PL)

1. The plant size should not exceed the one established by the platform and allow lateral side and top view as shown in Fig. 2. The resolution of digital images is 2454 × 2056 H x V pixels, with a pixel size of 4.4 × 4.4 μm²; this is obtained using a camera Basler AG, Ahrensburg (Germany), or equivalent.

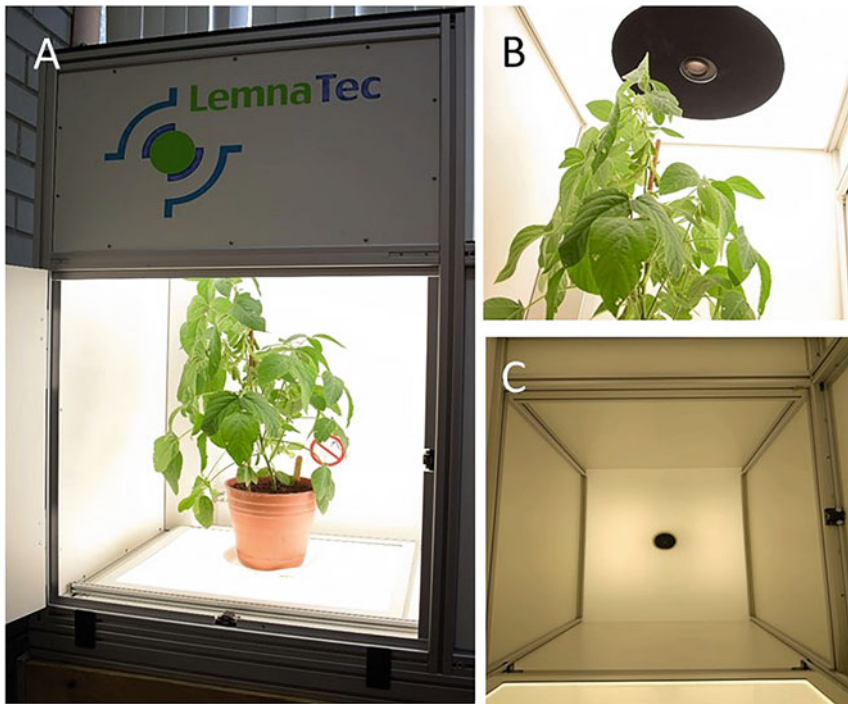


Fig. 1 The panel shows images of (a). Scanalyzer PL platform (LemnaTec, Germany). The cabinet is equipped with RGB imaging in (b), top, (c) side views

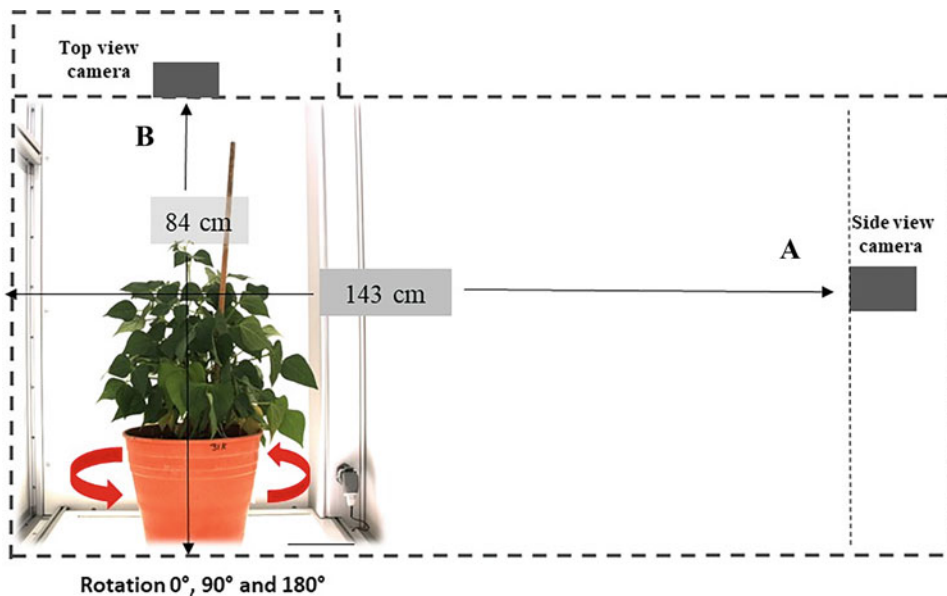


Fig. 2 The image shows reconstruction of field of views. (a) width = 54 cm, depth = 154 cm. (b) width = 54 cm, depth = 84 cm field of view, Min. Field of top view 4.6 cm × 6.1 cm. Max. 38 cm radius

2. Each plant is imaged individually in three (0° , 90° , and 180°) plane orientations, and monitored over course of time treatment or sampled progressively at the end of the cycle when the plant reached maturation (Fig. 2).
3. To guarantee a high-quality image, it is advised to select top or side camera with illumination settings for different applications. Each image configuration can store large numbers of configurations and recreate the exact same imaging conditions throughout the study.

3.2 Image Analysis

1. All components defining the image acquisition of a certain sensor (camera, optics, light, rotation) must be decided, and this configuration (LemnaControl) can be stored to define imaging conditions during the experimental by color classification cuantificating the damage and growth of each plant.
2. Once images are obtained (snapshots), the core database is analyzed with the commercial software, LemnaBase, LemnaGrid, and LemnaMine, to ensure the quality and lasting value of all acquired data and related to parameters as color classification cuantificating the damage and growth of each plant. Firstly, extract data from Lemna databases and employ user-friendly tools to organize and visualize them. By correlating image analysis data with other experimental data, the results are transformed into biologically relevant information for subsequent statistical analysis.
3. Database administration tools are provided with the LemnaMiner software suite. The databases will include the number of samples screened per time, test system or user, and control data throughout experiment.
4. Images are processed with the LemnaGrid software (Fig. 3). This software works as toolbox in a graphical dataflow programming language which allows easily connecting different algorithms, creating an image processing pipeline to extract the desired properties from the original image. To fulfill the needs of high-throughput image processing, the image processing chains (grids) are usually designed for representative images of the experiment and can then be applied to whole datasets consisting of thousands of images. In each image processing chain, there are four typical steps to extract properties from the image. (A) Load reference image. One or multiple reference images are loaded. These images should represent the most average and extreme cases to make the grid cover all possible images in the dataset that it is later applied. (B) Separate object from background. There are plenty of algorithms available to separate the foreground from the background, in addition to simple techniques like picking background color from the RGB

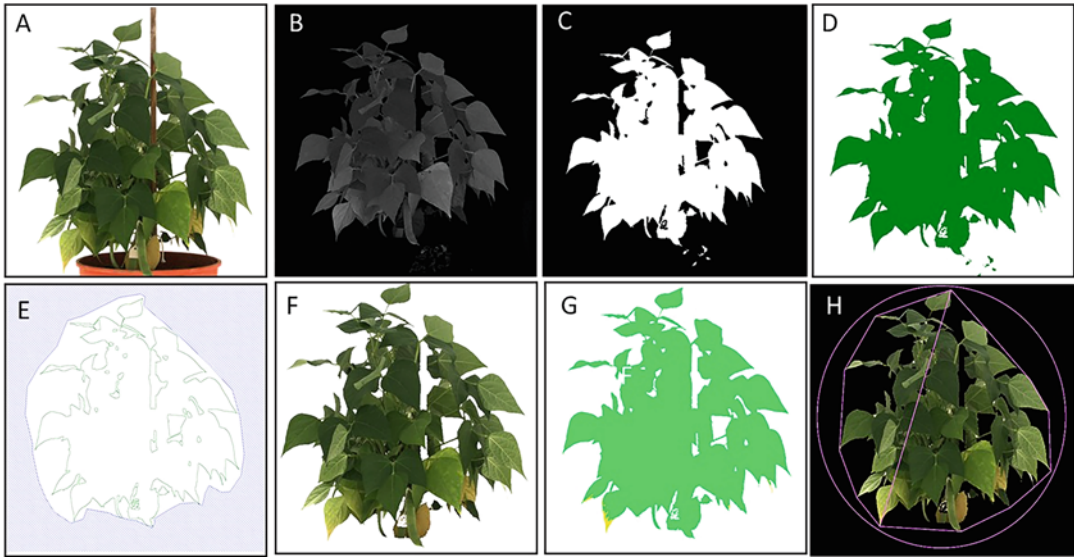


Fig. 3 Example of images taken from side views of common bean plants under drought and derivation of geometric parameters. First, the foreground and background are separated in the image (A–D), resulting in a binary image (E–G). Thereafter, the object is separated according to a color classification (highlighted in green, yellow senescence, and necrotic pink), and geometric parameters, such as caliper length, convex hull, and compactness as minimum enclosing area (H). The measurements are made in individual plants

color space. (C) Attach desired properties. As soon as the object is separated from the background, it contains a lot of mathematical properties such as size, length, width, and many more (Fig. 3H). Other parameters can be attached to the objects as well using additional algorithm such as color classification or skeleton information (Fig. 3G).

5. The size and dimensions of the object could be calculated, and all three images are used to estimate the overall biomass of the plant and compared with actual plant size determined on the destructively harvested plants. In addition, extracted plant pixel area from all side and top view images could be used to calculate a volume which was termed “digital biomass” that corresponds to a pixel volume using the following equation:

$$\text{Digital biomass average pixel side area top area} = * 2$$

6. This digital trait is used as a proxy for fresh weight. Traits estimated from side view images are plant height and plant width.
7. Additional types of automated imaging analysis include geometric parameters of convex hull (the smallest possible mathematically solved perimeter that envelopes the imaged plant), compactness (the ratio of leaf area per convex hull area), caliper length (the longest dimension of the canopy when viewed from above), circumference (the minimum circle that can enclose the

plant), and surface coverage (the ratio of leaf area to the area of the minimum enclosing circle calculated from the top view image). Compactness and convex hull measure the degree of leaf area coverage (Fig. 3A–F).

3.3 Statistical Analysis

Predicted values for genotype $\times$ experiment ($G \times E$) is calculated using the R function predict.asreml using the Excel procedure for linear regressions. Statistical analyses used for to test the effect of replications, genotypes, and stress treatment, the one-way analysis of variance (ANOVA) with subsequent post hoc pairwise comparison using Tukey Honest Significant Difference (HSD) could be applied at 95% confidence level. Pearson coefficients could be calculated to analyze the relevance of the stress treatment.

4 Notes

1. Additional biological replicates may be needed if the stress applied is too subtle.
2. In application to time-series images is only for plants with maximum dimensions mentioned in Fig. 2.
3. In order to exclude artifacts in the images, the pots must avoid reflections, so opaque colors other than green should be used (e.g., black) (Fig. 3A).
4. When the leaves overlapped, obtain as many snapshots as possible rotating the pot manually in different angles (Fig. 2).

Acknowledgments

We thank Cátedras-CONACyT program and Infrastructure Project CONACyT, INFRA-2015 (256307).

References

1. Abdelrahman M, El-Sayed M, Jogaiah et al (2017) The “STAY-GREEN” trait and phytohormone signaling networks in plants under heat stress. *Plant Cell Rep* 36:1009–1025
2. Gregersen PL, Culetic A, Boschian L et al (2013) Plant senescence and crop productivity. *Plant Molec Biol* 82:603–622
3. Kuska MT, Behmann J, Großkinsky DK et al (2018) Screening of barley resistance against powdery mildew by simultaneous high-throughput enzyme activity signature profiling and multispectral imaging. *Frontiers Plant Sci* 9: 1074–1074
4. Rebetzke GJ, Jimenez-Berni JA, Dovill WD et al (2016) High-throughput phenotyping technologies allow accurate selection of stay-green. *J Exp Bot* 67:4919–4924
5. Maxwell K, Johnson GN (2000) Chlorophyll fluorescence—a practical guide. *J Exp Bot* 51: 659–668
6. Jansen M, Gilmer F, Biskup B et al (2009) Simultaneous phenotyping of leaf growth and chlorophyll fluorescence via GROWSCREEN-FLUORO allows detection of stress tolerance in *Arabidopsis thaliana* and other rosette plants. *Funct Plant Biol* 36:902–914

7. Woo NS, Badger MR, Pogson BJ (2008) A rapid, non-invasive procedure for quantitative assessment of drought survival using chlorophyll fluorescence. *Plant Methods* 4:27–27
8. Thomas H, Ougham H (2014) The stay-green trait. *J Exp Bot* 65:3889–3900
9. Padilla-Chacón D, Peña Valdivia CB, García Esteva A et al (2019) Phenotypic variation and biomass partitioning during post-flowering in two common bean cultivars (*Phaseolus vulgaris* L.) under water restriction. *South African J Bot* 121:98–104



Chapter 2

An Automated High-Throughput Phenotyping System for *Marchantia polymorpha*

Karina Medina-Jimenez, Mario A. Arteaga-Vazquez, and Argelia Lorence

Abstract

High-throughput phenotyping (HTP) allows automation of fast and precise acquisition and analysis of digital images for the detection of key traits in real time. HTP improves characterization of the growth and development of plants in controlled environments in a nondestructive fashion. *Marchantia polymorpha* has emerged as a very attractive model for studying the evolution of the physiological, cellular, molecular, and developmental adaptations that enabled plants to conquer their terrestrial environments. The availability of the *M. polymorpha* genome in combination with a full set of functional genomic tools including genetic transformation, homologous recombination, and genome editing has allowed the inspection of its genome through forward and reverse genetics approaches. The increasing number of mutants has made it possible to perform informative genome-wide analyses to study the phenotypic consequences of gene inactivation. Here we present an HTP protocol for *M. polymorpha* that will aid current efforts to quantify numerous morphological parameters that can potentially reveal genotype-to-phenotype relationships and relevant connections between individual traits.

Key words High-throughput phenotyping, *Marchantia polymorpha*, Morphological parameters

1 Introduction

Even though early land plant evolution is still under intense debate [1], it is widely accepted that liverworts belong to the group of basal land plants (known as bryophytes, which also includes hornworts and mosses) that are closely related to the aquatic ancestor of land plants [2]. *Marchantia polymorpha* (*Marchantia*) is a dioecious liverwort with eight autosomes and either a sexual X or Y chromosome ($n = 9$) that has reemerged as an exciting model for evolution and functional genomics studies [3–5]. The life cycle of *Marchantia* involves a dominant haploid gametophytic phase characterized by the development of a dorsiventral thallus exhibiting a periodic bifurcated pattern [6, 7]. The thallus is considered the main plant body and it will produce distinct sets of tissues and organs. During the vegetative growth, the dorsal side will develop

air chambers and gemma cups that produce vegetative propagules termed gemmae (singular: gemma). The ventral side will develop two types of single-celled rhizoids (pegged and smooth) and scales [7]. The reproductive phase (that can be induced under controlled laboratory conditions with a combination of white and far-red light) is characterized by the development of either a male (in plants containing a Y chromosome) or a female (in plants containing an X chromosome) gametophore. The male antheridiophore, harbors antheridia that produce motile sperm (antherozoids), and the female archegoniophore, harbors archegonia within which egg cells develop [5]. Sexual reproduction results in the formation of a diploid embryo that will divide meiotically to produce thousands of haploid spores [reviewed in [7]]. When germinated, spores will subsequently produce a thallus, completing this way the sexual life cycle. The availability of in vitro culture techniques for rapid growth and propagation position *Marchantia* as a very powerful model for functional genomics as the sexual life cycle can be completed in less than 2 months and the vegetative one in less than 20 days. Additionally, protocols for chemical, physical, and biological mutagenesis (vgr. EMS, UV light, and T-DNA, respectively) are already available [5, 8].

Plant phenotyping focuses on the analysis of the interactions between the genome (and also the epigenome) and the environment (including both internal and external cues) and their impact on observable plant traits [9]. New technological advances such as high-throughput phenotyping systems (HTPS) are promising techniques for imaging and data processing that can be used to produce accurate measurements of the morphology and geometrical features of the plant through its life cycle [10]. Usage of HTPS in combination with functional genomics approaches will greatly aid current efforts to understand how gene functions shape plant growth and performance [11].

2 Materials

2.1 *Marchantia polymorpha* Gemmae

We employed gemmae from *M. polymorpha* plant accession Takaragaike-1 (Tak-1) that have been asexually maintained and propagated in vitro through asexual reproduction as previously described [12, 13].

2.2 Equipment and Materials

1. Graduated cylinders (1 L).
2. Flasks with sterile distilled water (2 L).
3. Graduated flask (1 L).
4. Graduated glass beaker .
5. Sterile petri dishes.

6. Sterile toothpicks.
7. Parafilm® to seal petri dishes.
8. 70% (v/v) ethanol.
9. Blue nylon mesh fabric discs (diameter 8 cm; *see* **Note 1**).
10. Scissors.
11. Sterile tweezers.
12. Analytical balance.
13. Magnetic stirrer.
14. Orbital shaker.
15. Laminar flow cabinet.
16. Vertical autoclave.
17. Culture room or growth chamber.
18. Autoclave.
19. Personal protective equipment (heat resistance gloves, latex gloves, and laboratory coat).

2.3 Culture Media

Gamborg B5 basal medium half strength (PhytoTech Labs), supplemented with 1% sucrose and 1% agar plant tissue culture (TC) grade (PhytoTech Labs). This media is the most commonly used for standard growth of gemmae in *M. polymorpha* (*see* Table 1).

As indicated, the culture media is prepared with purified deionized water (obtained from a purifying system with a sensitivity of 18 MΩ-cm at 25 °C). The media needs to be sterilized at 121 °C for at least 20 min using saturated steam under at least 15 pounds per square inch (psi) of pressure. The culture media and their components should be stored away from light, and exposure to direct sunlight should be avoided at all times. This protocol has been optimized to phenotype *M. polymorpha* thalli [12].

2.4 Marchantia Growth Conditions

Gemmae were grown at 22 °C under continuous white light (75 micromol m⁻² s⁻²) using a Conviron (Winnipeg, Canada) climate-controlled growth chamber.

2.5 Image Acquisition

To record the morphological parameters, a non-invasive image acquisition was carried out using the Scanalyzer HTS high-throughput phenotyping system (LemnaTec, Germany). The unit has an automatic robotic arm that holds three high-resolution cameras that allow top-down imaging of visible (VIS), fluorescence (FLUO), and the near-infrared (NIR) spectra. Visible images were taken with a piA2400-17gc CDD camera (Basler, Germany) with a resolution of 2454 × 2056 pixels and a sca1600-14gc CCD camera (Basler, Germany) with a resolution of 1624 × 1234 pixels for the FLUO images. The images were acquired in a sequential

Table 1
Preparation of Gamborg B5/2 media

Media compound	g/L
Gamborg B5 basal medium	1.5
Agar plant TC (1%)	10
Sucrose (1%)	10

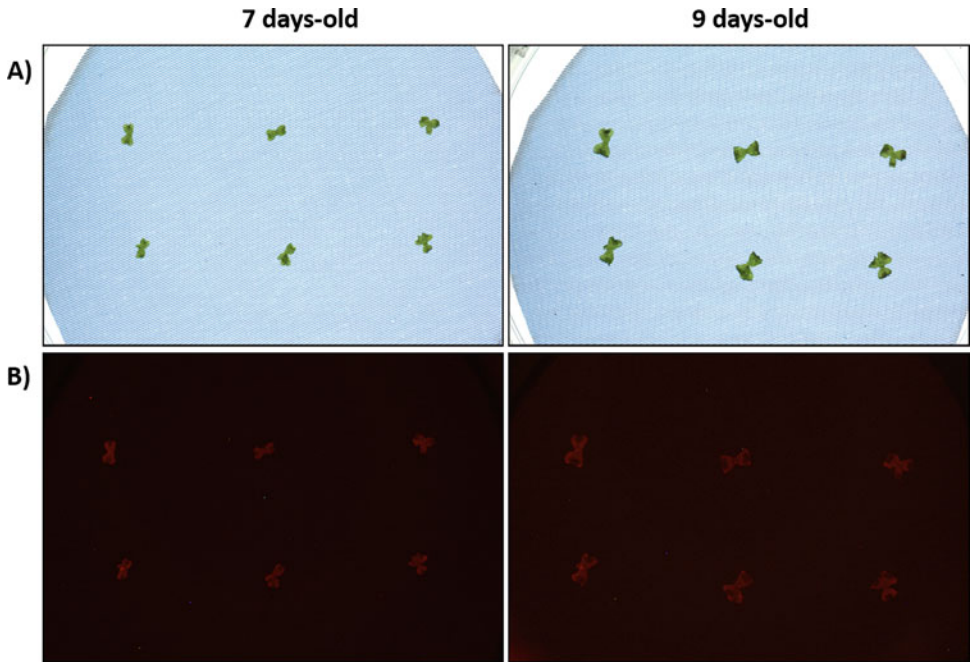


Fig. 1 Example of VIS and FLUO images with 7- and 9-day-old *M. polymorpha* thalli

manner in the same imaging station. The image files were exported in PNG (Fig. 1) and TIFF formats.

3 Methods

**3.1 Gamborg B5
Media Preparation**

1. For 1 L, weight 1.55 g of Gamborg basal salt mixture, 10 g of sucrose, and 10 g of agar plant tissue culture grade.
2. In a 2 L graduated beaker, add 500 mL of distilled water to dissolve Gamborg basal salt and sucrose using an orbital shaker and a magnetic stir bar (Table 1).
3. When the powder is completely dissolved, add 10 g of agar plant tissue culture (TC) grade. Calibrate the pH to 5.8 and adjust the volume to 1 L with sterile distilled water.

4. Autoclave at 121 °C for 20 min and after the cycle is complete, place the media into a laminar flow chamber to cool to 50 °C.
5. Set your plates and pour the Gamborg B5/2 medium into the petri dishes. Just pour enough media into the plates to cover the bottom of the plate. Every plate needs to be labeled with the date and the name of the *Marchantia* accession, genotype, treatment, or condition being analyzed.

3.2 Transferring Gemmae to the Culture Media

1. Once the Gamborg B5/2 media is solidified, place a blue nylon mesh fabric disc to cover the media using sterile tweezers (*see Note 1*).
2. Gemmae are directly collected from gemma cups of 3-week-old *M. polymorpha* thalli (*see Note 2*). The number of gemmae placed into the plates can be different, depending on the experiment. In this case, six gemmae were placed on each petri dish using a dissecting needle (heat sterilized) or sterile toothpicks (*see Note 3*).
3. In order to prevent contamination, petri dishes are sealed using Parafilm® or micropore.
4. Finally, the plates with gemmae are placed into the growth chamber, under the conditions previously described in the materials. Within 3–7 days, gemmae will develop into a thallus (*see Note 4*).

3.3 Imaging Acquisition

The imaging process started by transferring the petri dishes with the gemmae to the imaging system. Different decks for pots, petri dishes, and multiwell plates are available depending on the species of interest. In this case, the petri dish deck option was chosen. Once all petri dishes were appropriately placed, the LemnaControl software was accessed, and the option “biotest” chosen. Under “measurement series,” plate positions with the appropriate configuration and measurement series were selected. Finally, a name for the experiment was given before starting imaging. Imaging time varied depending on the number of plates to scan.

3.4 Image Analysis

Images from *Marchantia* gemmae were analyzed with the help of the image analysis software package LemnaBase (LemnaTec GmbH, Germany) (*see Note 5*). Images were processed using a user-designed pipeline for LemnaControl software. Before accessing the LemnaBase, and starting analysis, images were viewed and organized selecting “snapshot viewer.” To analyze the images, the option “image analysis” was selected (Fig. 2). Using visible images, we were able to obtain phenotypic parameters from each gemma including projected area (cm²), convex hull area (area that entirely encloses the plant) (cm²), caliper length (maximum diameter of the plant) (mm), and compactness (ratio of projected area to convex hull area). Using FLUO images, we were able to analyze the amount of chlorophyll present. The analyzed images and calculations were saved in the LemnaBase as a PostgreSQL database. The

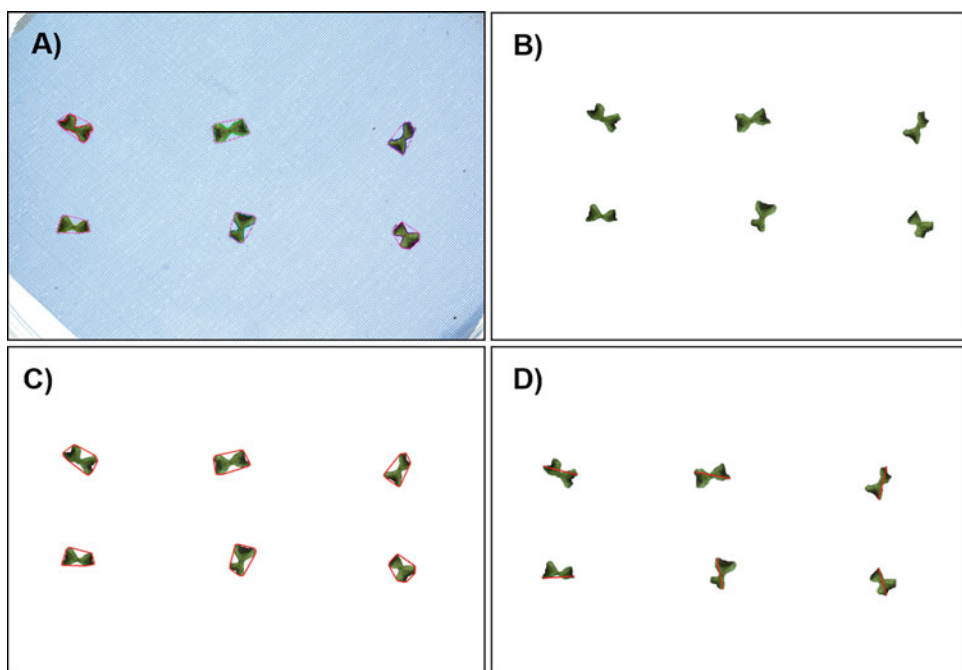


Fig. 2 Example of VIS images for 9-day-old thalli analyzed with the LemnaGrid software. (a) Identification of thalli from the background with the help of the blue mesh as part of the image processing. (b) Identification of the projected area. (c) Convex hull area. (d) Caliper length

output readouts from the images were exported as a CSV file for further statistical analyses.

4 Notes

1. In order to increase the contrast between the background and the region of interest (plant region) during image analysis using the software package LemnaGrid, a roughly uniform blue background (nylon blue mesh or something similar) must be used. Blue color is a good choice in terms of image quality [7]. Several circles made with blue nylon mesh fabric of the same size as the petri dishes were cut, autoclaved, and placed on the media.
2. The number of gemmae per plate will depend on the viability of gemmae. Within 3–5 days, the gemmae will become large thalli, not enough to overlap with each other. But if gemmae are older than 15 days, they will start overlapping.
3. Technical replicates of gemmae of the same age (coming from the same gemmae cup) were used for image acquisition. A size marker can be included next to the gemmae as a reference to convert the number of pixels to mm or cm.

4. One method involves using a paper disc with a grid, the same size as the petri dish to map the positions where gemmae will be placed. By placing the paper discs under the petri dishes as a reference, you are assuring that all of them are equally and consistently spaced.
5. The images can be analyzed using other image software tools (<https://www.quantitative-plant.org/software>), for example, PlantCV2 that is an open-source image analysis software package targeted for plant phenotyping (<https://plantcv.danforthcenter.org>).

Acknowledgments

The Scanalyzer HTS was acquired with funds from the Plant Powered Production (P3) Center by NSF grant # EPS-0701890. This work was funded by NSF EPSCoR Track 2 grants # IIA-1430427 and #1736192 to AL. We thank additional funding provided by the Arkansas Biosciences Institute and the Arkansas Research Alliance to AL. This work was supported by Consejo Nacional de Ciencia y Tecnología (CONACYT) grant A1-S-38383 and UCMEXUS-CONACYT grant CN-20-166 to MAAV.

References

1. Morris JL, Puttick MN, Clark JW et al (2018) The timescale of early land plant evolution. *Proc Natl Acad Sci U S A* 115(10): E2274–E2283
2. Bowman JL, Floyd SK, Sakakibara K (2007) Green genes comparative genomics of the green branch of life. *Cell* 129:229–234
3. Bowman JL (2015) A brief history of *Marchantia* from Greece to genomics. *Plant Cell Physiol* 57(2):210–229
4. Berger F, Bowman JL, Kohchi T (2016) *Marchantia*. *Curr Biol* 26(5):R186–R187
5. Ishizaki K, Ryuichi N, Yamato K T et al (2015) Molecular genetic tools and techniques for *Marchantia polymorpha* research. *Plant Cell Physiol* 57:262–270
6. Junker A, Muraya MM, Weigelt-Fischer K et al (2014) Optimizing experimental procedures for quantitative evaluation of crop performance in high throughput phenotyping systems. *Front Plant Sci* 5:770
7. Solly JE, Cunniffe NJ, Harrison CJ (2017) Regional growth rate differences specified by apical notch activities regulate liverwort thallus shape. *Curr Biol* 27(1):16–26
8. Takenaka M, Yamaoka S, Hanajiri T et al (2000) Direct transformation and plant regeneration of the haploid liverwort *Marchantia polymorpha* L. *Transgenic Res* 9(3):179–185
9. Shimamura M (2015) *Marchantia polymorpha*: taxonomy, phylogeny and morphology of a model system. *Plant Cell Physiol* 57(2): 230–256
10. Fiorani F, Ulrich S (2013) Future scenarios for plant phenotyping. *Annu Rev Plant Biol* 64: 267–291
11. Rahaman M, Chen D, Gillani Z et al (2015) Advanced phenotyping and phenotype data analysis for the study of plant growth and development. *Front Plant Sci* 6:619
12. Sozzani R, Benfey PN (2011) High-throughput phenotyping of multicellular organisms: finding the link between genotype and phenotype. *Genome Biol* 12(3):219
13. Bowman JL, Kohchi T, Yamato KT et al (2017) Insights into land plant evolution garnered from the *Marchantia polymorpha* genome. *Cell* 171(2):287–304



Chapter 3

A Novel High-Throughput Phenotyping Hydroponic System for Nitrogen Deficiency Studies in *Arabidopsis thaliana*

Lucia M. Acosta-Gamboa, Zachary C. Campbell, Fei Gao, Benjamin Babst, and Argelia Lorence

Abstract

High-throughput phenotyping enables the temporal detection of subtle changes in plant plasticity and adaptation to different conditions, such as nitrogen deficiency, in an accurate, nondestructive, and unbiased way. Here, we describe a protocol to assess the contribution of nitrogen addition or deprivation using an image-based system to analyze plant phenotype. Thousands of images can be captured throughout the life cycle of *Arabidopsis*, and those images can be used to quantify parameters such as plant growth (area, caliper length, diameter, etc.), *in planta* chlorophyll fluorescence, and *in planta* relative water content.

Key words Nitrogen, High-throughput phenotyping, Hydroponics, *Arabidopsis*

1 Introduction

Of all the plant nutrients, nitrogen (N) is the most important inorganic element necessary to support plant growth and development. Nitrogen is taken up in different forms, such as nitrate and ammonium, which can then be converted into the N component of amino acids, the main backbone of plant proteins, as well as numerous other crucial biochemicals including nucleotides, and the plant signal molecules auxin, cytokinin, and nitric oxide. This inorganic element also plays a role in the formation of secondary metabolites, chlorophyll, and coenzymes [1]. Plants require more N than any other mineral element, and N deficiency has been shown to limit plant growth as well as lower photosynthetic capacity and antioxidant production [2, 3]. *Arabidopsis* is typically grown in peat-based potting mix, which has a relatively high N content. In a hydroponics system, researchers have precise control over the amount of N introduced into the system, and this allows a better understanding of how much N plants need.

Traditionally, N-deficient phenotypes—and many other phenotypes observed under abiotic stress—have been measured manually with parameters such as height, fresh and dry weight of aerial and root tissue, yield, etc. These measurements can often be destructive, resulting in one-off measurements related to plant health and viability. However, high-throughput phenotyping is quickly revolutionizing the field, allowing plant phenotypes to be studied in greater depth than ever before. Utilizing a LemnaTec HTS Scanalyzer system, it is possible to monitor an experiment temporally from the seedling to reproductive stage in a high-throughput and nondestructive manner. In this work, we describe a high-throughput phenotyping method using *Arabidopsis thaliana*, a model plant system, to hydroponically study varying levels of N deficiency.

2 Materials

- Prepare all the solutions using ultrapure water (prepared by purifying deionized water, to attain a sensitivity of 18 M Ω -cm at 25 °C) and analytical grade reagents. Prepare and store all reagents at room temperature (unless otherwise specified). Diligently follow all waste disposal regulations when disposing of waste materials.
- 2.1 *Arabidopsis* Seeds** *Arabidopsis* wild-type Col-0 and T-DNA mutants (*nrt1.2*, *nrt1.6*, *nrt1.7*, *pot*) were obtained from the Arabidopsis Biological Resource Center (ABRC, Columbus, OH, USA).
- 2.2 Half MS Media** Seeds were germinated in ½ Murashige and Skoog (MS) medium [4], and grown for 10–12 days under a short-day photoperiod (10 h light, 24 °C daytime, 22 °C dark, 65% humidity). Light was provided by fluorescent tubes (F40PL/AQ/ECO 49893 40 W T12, GE) at a photon flux density of 150–200 $\mu\text{mol}/\text{m}^2/\text{s}$, measured at the top of the pots.
- 2.3 MS Salts** For this experiment, MS salts from Caisson Labs were used. This included Murashige and Skoog with Gamborg’s vitamins (MSP06) and Murashige and Skoog without nitrates (MSP07).
- 2.4 Quick Pot 15 Trays and Solid Matrix** Each genotype/treatment was grown in 85 × 73 mm Quick Pot 15 RW trays (HerkuPlast Kubern GmbH). The trays contained Profile Greens Grade Soil Amendment. For the half scoop of soil placed in the center of the solid matrix, *Arabidopsis* plant growing media was used (PRO-MIX PGX). Blue mesh (Kittrich Corporation, Pomona, CA) was placed on top of the soil mixture to prevent algal growth, reduce transpiration, and improve the object segmentation during image analysis (Fig. 1).



Fig. 1 Representative *Arabidopsis* images acquired with the vis (RGB, left), fluorescence (FLUO, center), and near-infrared (NIR, right) sensors

2.5 Image Acquisition and Analysis

Phenotyping was done using a Scanalyzer HTS high-throughput phenotyping system with the LemnaControl software (LemnaTec, Aachen, Germany) three times per week, starting 3 days after transplanting to monitor plants in the vegetative stage through the transition to the reproductive stage. The system has a robotic arm fitted with cameras that allow the capture of images, including visible (RGB), fluorescence (FLUO), and near-infrared (NIR). Differences between N treatments and genotypes are analyzed using LemnaGrid, and readouts such as rosette size, leaf shape and area, *in planta* chlorophyll fluorescence, and *in planta* water content are available. Image acquisition and analysis was done as previously described [5]. Phenotyping experiments were terminated when the flower stalks reached the camera, which affected the resolution of the images.

3 Methods

3.1 Seed Sterilization, Tissue Culture, and Vernalization

1. Sterilize: Place seeds in a 1.5 mL tube. Wash the seeds with 1 mL of 70% ethanol for 1 min. Spin the seeds down using a centrifuge ($4000 \times g$ for 1 min) and discard the ethanol. Wash the seeds with 1.5 mL of a solution of 50% bleach and 0.05% Tween 20 solution for 5 min. Spin the seeds again and discard the solution. Rinse the seeds six to eight times with sterilized deionized water until the bleach smell is gone. Perform all these steps using a laminar flow clean bench.
2. Plate: Using a flame-sterilized spatula, spread around 25–30 seeds in a petri dish containing $\frac{1}{2}$ MS phytagel medium. Seal the petri dish with Parafilm.
3. Vernalize: Place the plates at 4 °C for 2–3 days.

3.2 Growth Conditions in Plates

1. Place the petri dishes inside a growth chamber set up with a photoperiod of 10 h light/14 h dark, 22 °C for 10 days. The light intensity should be 150–200 $\mu\text{moles}/\text{m}^2/\text{s}$.

3.3 Nutrient Solution Preparation

1. Make the MS media for three N treatments by mixing MS and MS containing no N as presented in Table 1.

3.4 Profile Greens Grade Mixed with Nutrient Solution

1. Prepare three different types of nutrient solutions and mix them with the profile greens before putting the soil matrix into each well. Use 1.5 L of nutrient solution per tray. The profile green/nutrient solution mix should be very wet.
2. Place a round piece of tight mesh screen (0.25 mm opening or smaller) at the bottom of each well to keep the profile greens from pouring out.
3. Tamp the mix into the wells; around 174 g of dry profile greens is enough to fill a well.

3.5 Transfer and Establishment

1. Push a shallow hole into the mixture using a clean dibble or a gloved finger. Place half a scoop of regular peat-based potting mix in each well and plant half a tray for each genotype (*At* mutant and *At* Col-0 wild-type) for 3 trays (one tray for each of the three different nutrient treatments). Each tray should have seven mutants and eight Col-0.
2. Using forceps, gently pull the 10-day-old seedlings out from the $\frac{1}{2}$ MS without breaking the roots.
3. Place the plant into the hole and gently push potting mix to cover the roots (*see Note 1*).
4. The profile greens should be wet enough, but if it looks dry, add a little water to the tray.
5. Cover the tray with a loose-fitting plastic lid for 2 days to keep moisture and ensure high humidity during establishment.
6. Remove the plastic lid and water the plants from the bottom once a week by filling the drainage tray in the morning and leaving the plants in standing water for 3–5 min, emptying the water once the profile greens surface looks wet.

Add nutrient solution once a week. In each N treatment, add the appropriate solution from the bottom (1.5 L/tray). During week 1, water once and fertilize once (*see Note 2*).

3.6 Phenotyping

1. Acquire images from plants every other day. Start scanning a day after transplanting. Obtain images using the visible (RGB), fluorescence, and near-infrared cameras to observe differences between N treatments and genotypes.
2. After completing the phenotyping, harvest the plant tissue as required.
3. Analyze images as described in [5, 6] (*see Note 3*).

Table 1
Nitrogen calculations for three different treatments

Nutrition treatment	[N] (mM)	No N MS: MS	For 1 L solution		For 1.5 L solution (100 mL/pot * 15 pots/ tray)	
			No N MS	MS	No N MS	MS
½ N	30	1:1	1.615 g	2.22 g	2.42 g	3.33 g
1/10 N	6	9:1	2.907 g	0.444 g	4.36 g	0.666 g
1/100 N	0.6	99:1	3.198 g	0.044 g	4.80 g	0.067 g

No N MS = Murashige and Skoog without nitrates

No need to autoclave if watering immediately after dissolving MS powders in water

4 Notes

1. Need to be very careful to transfer the seedlings; mortality rate is higher than in regular soil. It is suggested to transfer extra backup seedlings for each well, wait for a few days, and if both of them survive, remove the least robust seedling.
2. For example, if plants were transferred on Tuesday, they should be watered Friday and fertilized with 1.5 L of nutrient solution on next Tuesday.
3. Alternatively, images can be analyzed with open-source software. We have successfully used PlantCV2 [7].

Acknowledgments

The Scanalyzer HTS was acquired with funds from the Plant Powered Production (P3) Center by NSF grant # EPS-0701890. This work was funded by NSF EPSCoR Track 2 grants # IIA-1430427 and #1736192 to AL. We thank additional funding provided by the Arkansas Biosciences Institute (ABI) and the Arkansas Research Alliance (ARA).

References

1. Pilbeam DJ (2011) The utilization of nitrogen by plants: a whole plant perspective. In: Foyer CH, Zhang H (eds) Annual plant reviews, nitrogen metabolism in plants in the post-genomic era. Wiley, New York
2. Huang ZA, Jiang DA, Sun JW et al (2004) Effects of nitrogen deficiency on gas exchange, chlorophyll fluorescence and antioxidant enzymes in leaves of rice plants. *Photosynthetica* 42(3):357–364
3. Kusano M, Fukushima A, Redestig H et al (2011) Metabolic approaches toward understanding nitrogen metabolism in plants. *J Exp Bot* 62(4):1439–1453
4. Murashige T, Skoog FA (1962) A revised medium for rapid growth and bioassays with tobacco tissue culture. *Physiol Plant* 15:473–497
5. Acosta-Gamboa LM, Liu S, Langley E et al (2017) Moderate to severe water limitation

- differentially affects the phenome and ionome of *Arabidopsis*. *Funct Plant Biol* 44:94–106. <https://doi.org/10.1071/FP16172>
6. Babst BA, Gao F, Acosta-Gamboa LM et al (2019) Three *NPF* genes in *Arabidopsis* are necessary for normal nitrogen cycling under low nitrogen stress. *Plant Physiol Biochem* 143:1–10
 7. Gehan MA, Fahlgren N, Abbasi A et al (2017) PlantCV v2.0: image analysis software for high-throughput plant phenotyping. *PeerJ* 5:e4088. <https://doi.org/10.7717/peerj.4088>



Chapter 4

***Camelina sativa* High-Throughput Phenotyping Under Normal and Salt Conditions Using a Plant Phenomics Platform**

Emilio Vello, John Aguirre, Yang Shao, and Thomas Bureau

Abstract

Climate change and environmental pollution will have a great impact on food security worldwide. More than 30% of the world's irrigated areas are estimated to be perturbed by high salinity affecting the productivity of crops. *Camelina sativa*, also known as false flax, is a flowering plant that is mainly cultivated as an oilseed crop that has many potential economic benefits; it can be used in food products, in industrial applications, and in animal feed and converted into biofuel. However, natural disasters due to climate events have led to significant crop losses. In this work, we developed a high-throughput phenotyping protocol to analyze the effects of different concentrations of salt on *C. sativa* using the McGill Plant Phenomics Platform (MP3). We present an adapted protocol to be applied with phenomics facilities in a greenhouse environment and the most effective way for high-throughput phenotyping.

Key words Phenomics, Phenotype, *Camelina sativa*, Salt tolerance, Abiotic stress, Image analysis

1 Introduction

In the last decade, plant phenomics has become essential to improve crop yield production and stress resistance [1, 2]. Traditional phenotyping methods in which plant harvesting is necessary are time and labor consuming, and analyses of growth dynamics on individual plants cannot be achieved [1]. Plant image-based phenotyping, therefore, is receiving more interest resulting in the increasing number of plant phenomics platforms and centers appearing around the world such as the McGill Plant Phenomics Platform (MP3, <http://mp3.biol.mcgill.ca>). These facilities allow researchers to accumulate hundreds or thousands of images under different portions of the light spectrum including other complementary sensors such as height detectors. These technologies have the advantage of collecting multidimensional high-throughput data quickly and in a noninvasive way, unthinkable with the

“old-fashion” methods [1, 3]. Here, we present a protocol to assess *C. sativa* varieties or mutant lines responses under different levels of salt concentration using a plant phenomics platform.

2 Materials

2.1 Plant Growth

1. Plastic round planter pots (diameter, 5"/12.7 cm; color, green) and plastic saucers according to the number of samples (*see Note 1*).
2. Pro-Mix BX, use 250 g of per pot.
3. Greenhouse room (semi-controlled environment) divided in growth zones (*see initial preparation below*).
4. *Camelina sativa* variety seeds: “Celine.”

2.2 Phenotyping Equipment

1. LemnaTec Scanalyzer conveyer system (3D); LemnaTec GmbH, Wuerselen, (Germany), equipped with multiple sensors: two visible light cameras piA2400-17gc (VIS), 2454 × 2056 pixels; two near-infrared cameras NIR-300PGE (NIR), 320 × 254 pixels; and two infrared cameras IRC-320GE (IR), 320 × 240 pixels (*Fig. 1*).
2. A total of 17 carriers for the 3D with a center hole to accommodate 8" round plastic pots (*Fig. 2*).
3. A total of 17 plastic saucers with a hole in the middle to accommodate 5" round plastic pots (*Fig. 2*).

2.3 Image and Statistical Software

1. In our case, a custom image analysis algorithm was developed using Java 1.8.0–45 (<http://www.java.com>) and ImageJ library v1.49t (<http://imagej.nih.gov/ij/>) [4]. The statistical analysis script was written in R v3.0.2 15 (<http://www.r-project.org/>). PostgreSQL v 9.3.1 was used to build the database. However, there are other image libraries, programming languages, and databases under open-source and commercial licenses that can be used to implement the image and data analysis algorithms and pipelines.
2. A Dell R910 server with 512 GB of RAM and two MD1200 storage devices 72 TB to process the image data analysis.

3 Methods

3.1 Initial Preparation

1. Fill up 64 pots (round 5") with 250 g of Pro-Mix BX. The number of pots depends on the experimental design.
2. Water the pots with 450 mL of water. Soil and water should weigh 700 g.

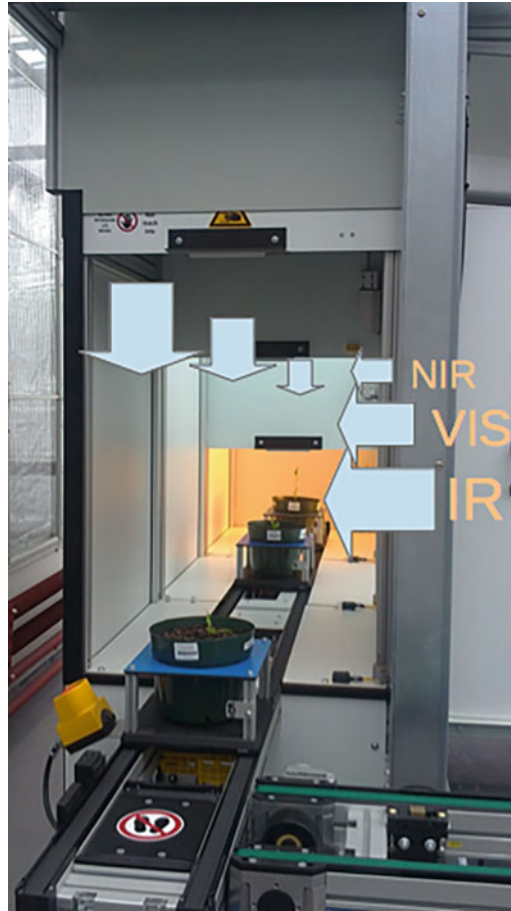


Fig. 1 Customized version of the LemnaTec 3D Scanalyzer installed in the McGill Plant Phenomics Platform (MP3 – <http://mp3.biol.mcgill.ca>). The system is equipped with two infrared cameras IRC-320GE (IR) (top-side), 320×240 pixels (first cabinet); two visible light cameras piA2400-17gc (VIS) (top-side), 2454×2056 pixels (second cabinet); and two near-infrared cameras NIR-300PGE (NIR) (top-side), 320×254 pixels (third cabinet)

3. Seed three *Camelina* seeds per pot (*see* **Note 2**).
4. Set up a greenhouse environment to 14 h:10 h light/dark photoperiod cycle. Temperature 24 °C day and 20 °C at night. Humidity between 40% and 60%.
5. Divide the samples into groups according to the different treatments (*see* **Note 3**).
6. Divide the greenhouse room into zones and assign each pot to a zone according to the number of samples. We don't assign more than 17 pots per zone. However, a minimum of two samples of each treatment group is assigned to each zone group (*see* **Note 4**).



Fig. 2 A *Camelina* plant in the visible light cabinet. A round 5" pot is located inside a homemade adapter from plastic saucers that is installed into 8" pots to fit the 3D carriers. The blue type around the bigger pot is to improve the image analysis process

7. Identify each pot with a label containing number (unique per pot), barcode, greenhouse room location, and treatment.

3.2 Growing Conditions

1. Thin seedlings after 1 week of growth to obtain one plant per pot. Choose the seedlings presenting similar size.
2. Every 2 or 3 days, weigh the pots and water to reach 700 g if needed. This operation can be performed automatically if the system is equipped with a watering system (*see Note 5*) (Fig. 3).
3. Every 3 days, randomly shift the position of the zone groups in the greenhouse room to avoid the position effect.

3.3 Salt Treatment

Salt stress is applied when plants have four pairs of fully developed leaves. This occurs at approximately 20 days after sowing (DAS).

1. Prepare NaCl solutions to 0, 50, 100, 150, 200, 250, 300, and 350 mM accordingly to a final volume of 450 mL of water. The concentration should be chosen according to the experimental design. However, a set of non-treated samples should always be part of the experimental design (Fig. 4).
2. Apply the treatment according to the sample identification label. The desired concentration is reached after four applications over 2 days (*see Note 6*).
3. Monitor the weights of the plants every day to keep 450 mL of water per pot (*see Note 7*).

3.4 Image Acquisition

A customized version of the LemnaTec Scanalyzer conveyer system (3D); LemnaTec GmbH, Wuerselen (Germany), installed at the McGill Plant Phenomics Platform (MP3 – <http://mp3.biol.mcgill.ca>) is used to carry out the image acquisition. However, an adapted version of this protocol can be implemented into any image-based phenotyping equipment if the system is able to accommodate plants of about 60 cm width and 90–100 cm height (*see Note 8*) (Figs. 1 and 3).

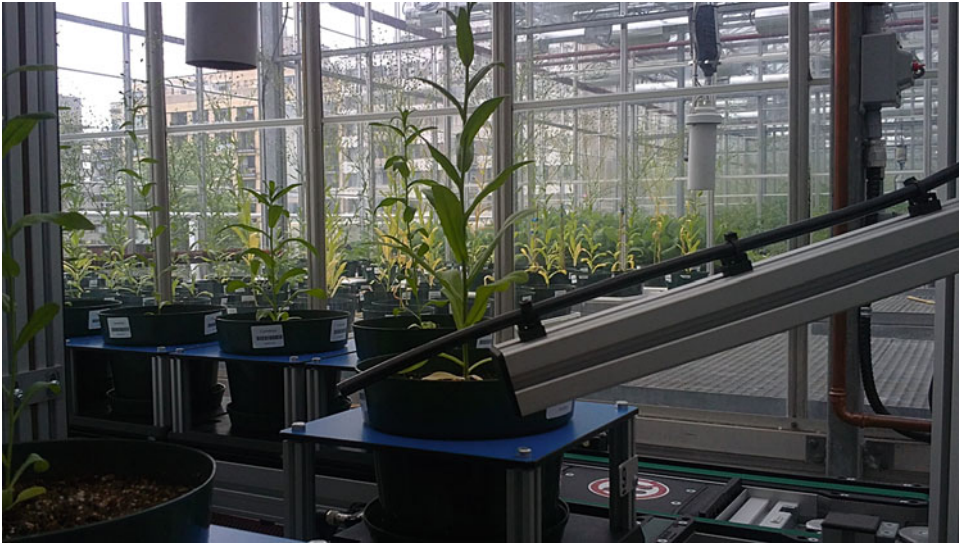


Fig. 3 Plants in the imaging cycle, watering station, and greenhouse. Plants in the conveyer system passing for the watering station. Samples are identified with a barcode to reduce manipulation errors

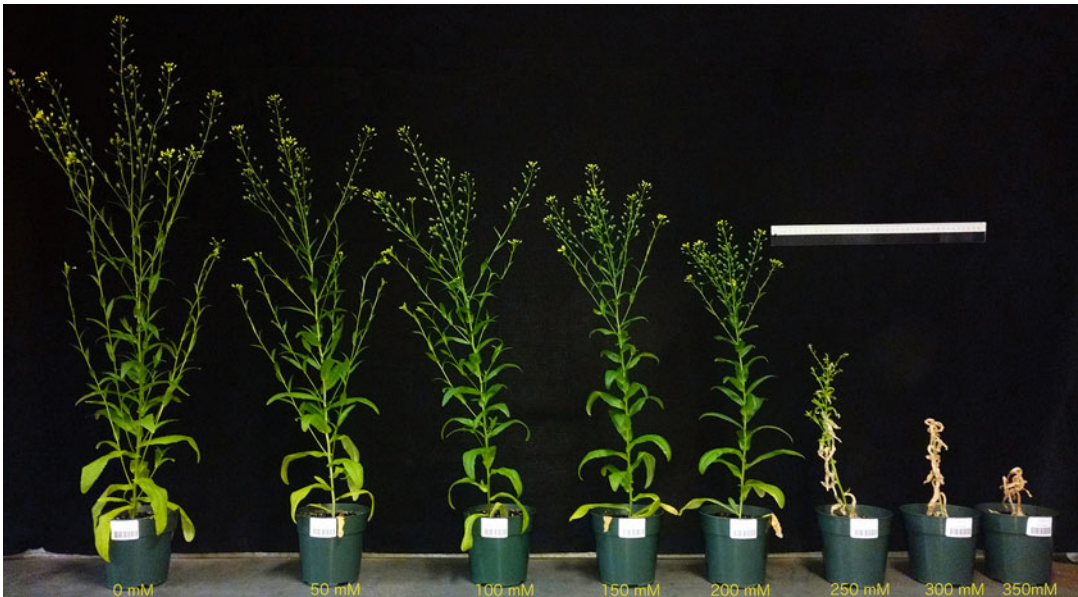


Fig. 4 A sample from each group treated with non-salt, 50, 100, 150, 200, 250, 300, and 350 mM, respectively, at 37 days after sowing

Table 1
Optical settings and image processing parameters of views

View	Light	Angle	Exposure	Zoom	Resize	x-offset	y-offset
vis-top-1-1000	VIS	0	80	1000	–	–	–
vis-top-min1-1	VIS	0	80	1	–	–	–
vis-side-1-0	VIS	0	100	1	–	–	–
vis-side-1-0-dk	VIS	0	40	1	–	–	–
vis-side-1-90	VIS	90	100	1	–	–	–
nir-side-1-0	NIR	0	120	1	298	8	5
nir-side-1-90	NIR	90	120	1	298	7	5
nir-top-1-1000	NIR	0	100	1000	298	10	2
nir-top-min1-1	NIR	0	100	1	298	13	0
ir-side-1-0	IR	0	–	–	375	-30	-75
ir-side-1-90	IR	90	–	–	375	-30	-75
ir-top-1-0	IR	0	–	–	290	13	2

1. Set up the camera configurations to be used during the experiment (Table 1). In our case, we have used a different combination of top and side views, rotation angles, and optical parameters. In this report, the word “view” is used to name each of those combinations (*see* Note 9).
2. Image every plant at regular time intervals. The desirable frequency is four times a week. However, a minimum of two measurements a week is necessary. In the 3D, pots are moved to the imaging cycle in blocks of 17 using the exchange option of the software.

3.5 Image Analysis

In addition to specialized software, databases, and powerful servers, the implementation of this step requires prior knowledge of image analysis and programming. This is especially important in high-throughput image-based phenotyping. We provide enough detail to allow computer developers to implement this pipeline in any programming language using most of the image analysis software libraries (*see* Note 10) (Fig. 5).

3.5.1 “vis-side-1-0” and “vis-side-1-90” Views

1. Convert the image to HSB color space.
2. Retain pixels having hue (H) channel value higher than 25 and lower than 103, saturation (S) higher than 30, and brightness (B) lower than 170.
3. Mark background as RGB (255, 255, 255) which represents the white color.

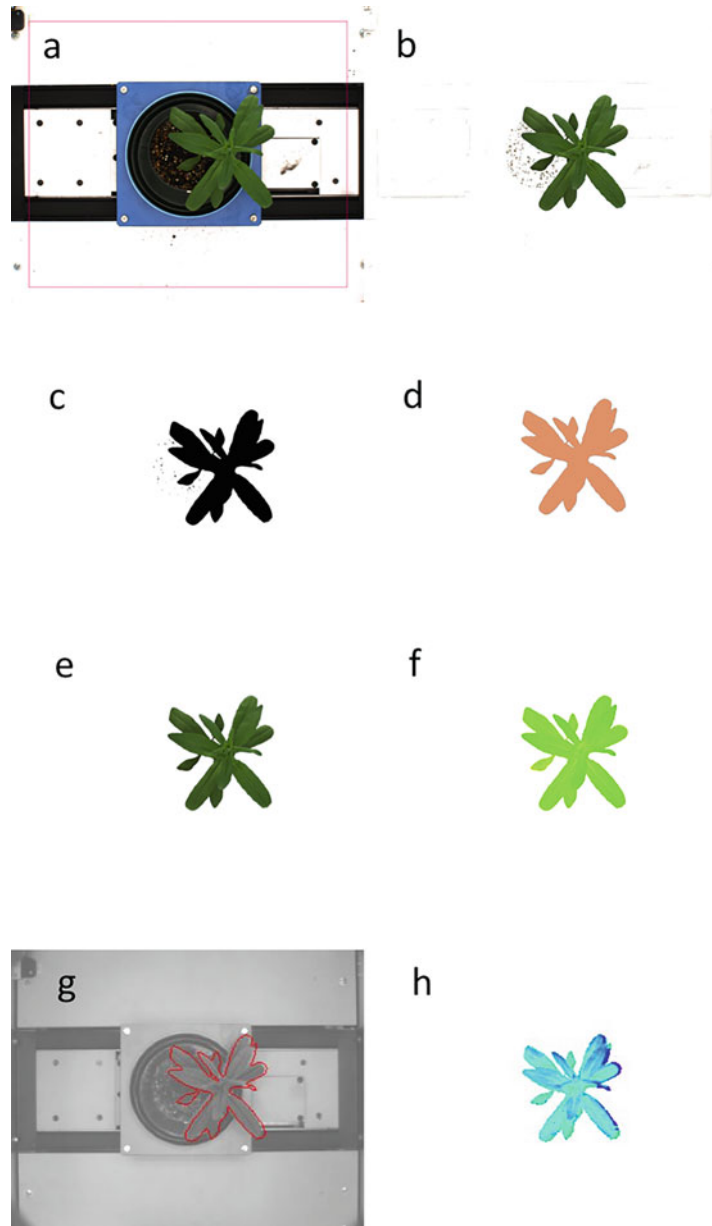


Fig. 5 Image analysis pipeline. Visual representation of some steps of the “vis-top-1-1000” view image process algorithms. **(a)** Original image including the virtual rectangle representing the interesting region. **(b)** Pixels retained after applying the following filters: (1) red (R), green (G), and blue (B) channel values lower than 198, (2) hue (H) channel value higher than 5 and lower than 103 and saturation (S) higher than 70, and (3) formula I result higher than 0.85 or a formula II result higher than 19. I) $(R-B) / (R + B)$. II) $4 \cdot G - 4 \cdot B - R$ [8]. **(c)** Pixels selected after applying two eroding and one dilating operations in a gray scale-type version of the image. **(d)** Pixels retained after the selection of objects with area higher than 700, a maximum distance to the theoretical image center of 500, and a maximum eccentricity value of 9. **(e)** Pixel original values of the region of interest or “digital plant”. **(f)** Reduction of the HUE channel of the HSB

4. After reducing image into gray scale type, submit the resulted image to eroding and dilating operations and reconvert into color image.
5. Tag the selected pixels as foreground.
6. Identify the objects based on the “combined contour tracing and region labeling” algorithm proposed by Burger and Burge [5, 6] (*see Note 11*).
7. Select objects: lying outside of the rectangles described by the coordinates (153, 660, 237, 2037) and (2247, 660, 2349, 2037), having an area greater than 500, a Euclidean distance lower than 1100 to the theoretical pot center, and an eccentricity lower than 80.
8. Join the resulted objects. Each plant will be represented by one object or “digital plant” from which the color-morphological features are calculated [3–5, 7].

3.5.2 “vis-top-1-1000” View

1. Select pixels having red (R), green (G), and blue (B) channel values lower than 198.
2. Convert resulted image into HSB color space.
3. Select pixels having hue (H) channel value higher than 5 and lower than 103 and saturation (S) higher than 70.
4. Apply the following formulas:
 - (a) I) $(R-B)/(R+B)$
 - (b) II) $4 \cdot G - 4 \cdot B - R$ [8]
5. Retain pixels having a formula I result higher than 0.85 or a formula II result higher than 19.
6. The rest of the algorithm is similar to Subheading 3.5.1 with some differences in the parameters.
 - (a) Perform two eroding and one dilating operations.
 - (b) Select objects with area higher than 700, a maximum distance to the theoretical image center of 500, and a maximum eccentricity value of 9.

3.5.3 “vis-side-1-0-0dk” View

The general structure of this algorithm is similar to Subheading 3.5.1.

1. Select pixels having saturation (S) higher than 10 and blue channel value higher than 85.
2. Did not apply formulas I or II and erosion-dilation operations.



Fig. 5 (continued) color space into color classes. (g) Near-infrared image with the mask from the visible light image in red. (h) False color representation of the near-infrared intensity of the “digital plant”

3. In this case, the minimum area is 50, the maximum distance to the theoretical center is 800, and the maximum eccentricity value is 100.

3.5.4 “vis-top-min1-1” View

The general structure of the Subheading 3.5.1 algorithm is followed.

1. Transform images into HSB color space.
2. Select pixels having hue (H) channel value higher than 25 and lower than 103 and brightness (B) lower than 220.
3. After converting into gray scale images, perform erosion and dilation operation to the images.
4. Convert the images back into color scale using the pixel information of the original images.
5. Join objects having an area greater than 700, a distance to a theoretical center bigger than 1000, and eccentricity value lower than 500 into one object.

3.5.5 Near-Infrared and Infrared Views

1. Use the corresponding images already treated with the visible light algorithms as masks or templates to obtain the plant pixels of the near-infrared and infrared images. A process of resizing and horizontal and vertical offsetting is necessary.
 - (a) “nir-side-1-0” view. Resize, 298; x-offset, 8; y-offset, 5
 - (b) “nir-side-1-90” view. Resize, 298; x-offset, 7; y-offset, 5
 - (c) “nir-top-1-1000” view. Resize, 298; x-offset, 10; y-offset, 2
 - (d) “nir-top-min1-1” view. Resize, 298; x-offset, 13; y-offset, 0
 - (e) “ir-side-1-0” and “ir-side-1-90” views. Resize, 375; x-offset, -30; y-offset, -75
 - (f) “ir-top-1-0” view. Resize, 290; x-offset, 13; y-offset, 2

3.5.6 Morpho-colorimetric Features

Morpho-colorimetric features are calculated from the digital plant as part of the image analysis algorithms. These features are used to assess differences among samples and treatment groups (*see Note 12*).

1. Area: number of pixels of the digital plant (Fig. 6).
2. Perimeter: length of the outer contour of the digital plant. In 8-neighborhoods, assign 1 to the horizontal and vertical segments and $\sqrt{2}$ to diagonal segments. Apply a 0.95 correction to the total value [5, 6].
3. Circularity: ratio between the circumference square and the area [7].
4. Compactness: ratio between the area and the perimeter [5, 6].

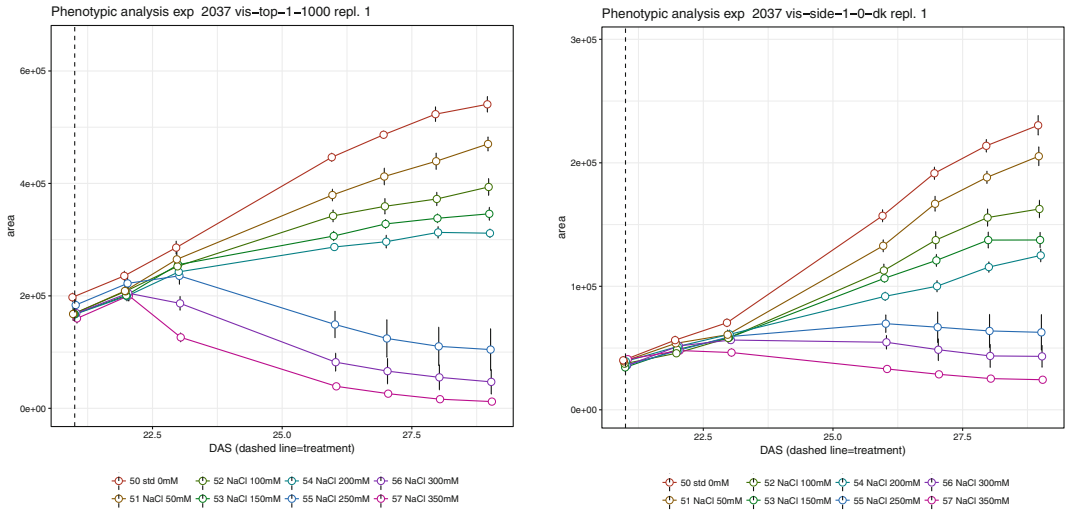


Fig. 6 Projected shoot area. Number of pixels of the area (mean $\pm$ standard error) as a function of the number of days after sowing (DAS). (a) “vis-top-1-1000”, (b) “vis-side-1-0-dk”

5. Major axis: axis where a physical body requires less effort to rotate. It extends from the centroid (center of gravity) to the widest part of the object [5, 6] in this case the “digital plant”.
6. Angle: orientation of the major axis [5, 6].
7. Minor axis: axis perpendicular bisector to the major axis.
8. Eccentricity: ratio between the major axis and the minor axis of the digital plant [5, 6]. The minor axis extends from the centroid to the narrowest part perpendicular to the major axis.
9. Color markers: first, second, and third quartile values from the pixel distribution of each particular RGB color space channel and gray scale $(R + G + B)/3$.

3.5.7 Color Classification and Clustering

The color classification and clustering approach is based on previously reported methods [2, 3].

1. Divide equally the hue channel of the HSB color space from visible light views into 64 categories. It is possible to use 32 and 16. Each class or category is defined as an interval of intensities.
2. Classify each pixel of the “digital plant” into one category.
3. Calculate a Euclidean distance matrix of the “digital plants” using the color classes.
4. Perform a hierarchical cluster analysis using the method “ward” of the R function “hclust” on the distance matrix.
5. Divide the resulting cluster tree of samples into two groups using the R function “cutree.”

3.6 Data and Statistical Analysis

1. The morpho-colorimetric features can be analyzed using any “standard” statistical methodology such as ANOVA or a linear mixed model which may be a more accurate way to assess time series in biological data [9] presented in most of the statistical software.
2. Color classification analysis: Build a 2×2 contingency table for each of the treated groups and the non-treated group as columns and the clustered groups as row. A Fisher’s exact (“fisher.test” function in R), Cochran-Mantel-Haenszel (“mantelhaen.test” function in R), or Pearson’s chi-squared test as a goodness-of-fit test (“chisq.test” function in R) can be applied to detect differences between groups at different days of sowing.

4 Notes

1. It is possible to use other pot size. This protocol has also been used with 8"/20.32 cm plastic planter pots.
2. We found that three seeds per pot is the optimal number to obtain at the end one plant per container. However, mutant line seeds are quite limited or may have a low germination rate. So, the number of seeds may vary accordingly.
3. In this paper, treatment groups are formed by different salt concentrations. However, these groups can be different mutant lines or a combination of concentration-mutant lines. This protocol has also been used to test overexpression mutant lines.
4. We have chosen 17 pots per zone because it is the maximum capacity of the image cycle in the McGill 3D system, and we haven’t observed any position effect [10]. However, this protocol has also been tested with eight plants per group. We think that any number below 17 will work well.
5. The 3D system is equipped with a watering system. If this option is chosen, the average weights of the carriers need to be considered. In the case of the MP3, the average carrier weight is about 1.935 kg.
6. The salt application is differed over time to avoid a plant “shock”. However, it depends on the experimental design. This protocol has also been used with “one shot” application.
7. This ensures that the salt concentration is constant over time.
8. In small systems, plants may be monitored to earlier stages. In high concentration of salt, the effect of the stress can be seen 1 to 4 days after the treatment.
9. The combination of views (cameras/angles/zooms/etc.) maximizes the collection of data from one single experiment.

Multiple views permit to validate the results as well as future analyses.

10. The algorithm and pipeline parameters need to be adjusted according to the phenotyping equipment and the imaging setup. However, the general structure of these algorithms will be valid in most of the cases.
11. Any other object detection algorithm could be used in this step.
12. Additional morpho-colorimetric features can be calculated from the “digital plant” such as width, height, median, or skew.

Acknowledgments

We would like to thank A Dao, M Cho, P Leroy, G Sahin, and J Tong Yang for helping with the phenotyping. We also thank M Romer and C Cooney for their help in the greenhouse and N Hernandez and M Letourneau for reviewing this manuscript. This project was funded by grants from the Canada Foundation for Innovation (CFI) and the Natural Sciences and Engineering Research Council (NSERC) of Canada to T Bureau.

References

1. Valle B, Simonneau T, Boulord R et al (2017) PYM: a new, affordable, image-based method using a Raspberry Pi to phenotype plant leaf area in a wide diversity of environments. *Plant Methods* 13(1):98
2. Joly-Lopez Z, Forczek E, Vello E et al (2017) Abiotic stress phenotypes are associated with conserved genes derived from transposable elements. *Front Plant Sci* 28. <https://doi.org/10.3389/fpls.2017.02027>
3. Vello E, Tomita A, Diallo AO et al (2015) A comprehensive approach to assess *Arabidopsis* survival phenotype in water-limited condition using a non-invasive high-throughput phenomics platform. *Front Plant Sci* 6:1–10
4. Schneider CA, Rasband WS, Eliceiri KW (2012) NIH image to ImageJ: 25 years of image analysis. *Nat Methods* 9(7):671–675
5. Burger W, Burge MJ (2008) Digital image processing. An algorithmic introduction using Java. First. Springer, New York
6. Burger W, Burge MJ (2016) Digital Image Processing. In: Gries D, Schneider F (eds) Algorithmic introduction using Java, 2nd edn. Springer-Verlag, London
7. Camargo A, Papadopoulou D, Spyropoulou Z et al (2014) Objective definition of rosette shape variation using a combined computer vision and data mining approach. *PLoS One* 9(5):e96889
8. Arvidsson S, Perez-Rodriguez P, Mueller-Roeber B (2011) Methods A growth phenotyping pipeline for *Arabidopsis thaliana* integrating image analysis and rosette area modeling for robust quantification of genotype effects. *New Phytol* 1(191):895–907
9. Poiré R, Chochois V, Sirault XRR et al (2014) Digital imaging approaches for phenotyping whole plant nitrogen and phosphorus response in *Brachypodium distachyon*. *J Integr Plant Biol* 56(8):781–796
10. Brien CJ, Berger B, Rabie H et al (2013) Accounting for variation in designing greenhouse experiments with special reference to greenhouses containing plants on conveyor systems. *Plant Methods* 9(1):5



Chapter 5

A Straightforward High-Throughput Aboveground Phenotyping Platform for Small- to Medium-Sized Plants

Denise Caldwell and Anjali S. Iyer-Pascuzzi

Abstract

High-throughput phenotyping platforms for growth chamber and greenhouse-grown plants enable non-destructive, automated measurements of plant traits including shape, aboveground architecture, length, and biomass over time. However, to establish these platforms, many of these methods require expensive equipment or phenotyping expertise. Here we present a relatively inexpensive and simple phenotyping method for imaging hundreds of small- to medium-sized growth chamber or greenhouse-grown plants with a digital camera. Using this method, we image hundreds of tomato plants in 1 day.

Key words Phenotyping, Aboveground, Digital camera, RGB images

1 Introduction

A growing population, increasing climate challenges, or decreasing arable land is driving the need for improved crop varieties with increased yield and stress tolerance. Understanding the relationship between plant genotype and phenotype is key to crop improvement [1, 2].

Technological advances have led to lower costs and increased efficiencies in genotyping, but high-throughput inexpensive methods of plant phenotyping remain a bottleneck to crop improvement [2, 3].

Recent years have seen an explosion of phenotyping platforms, software, and resources for nondestructively phenotyping and analyzing a vast range of above- and below-ground plant traits [1, 2, 4–8]. Phenotyping platforms are available for plants grown at multiple scales—from growth chamber to greenhouse to the field. With available technology, scientists can measure nearly any above-ground trait they wish. For example, using thermal infrared cameras, leaf temperature can be recorded. Near-infrared cameras can provide data regarding leaf water content, and hyperspectral

imaging can provide researchers with spectral signatures of abiotic and biotic stress responses [4–6]. Digital cameras, which image in the visible light range, result in images that can be analyzed for data regarding plant architecture, growth rates, shape, length, color, and biomass. These advances have added tremendously to the plant phenotyping field, and with improved image processing algorithms, have made it possible to identify phenotype-genotype linkages. However, most of these platforms are very expensive, and require substantial investments in time and expertise to acquire and establish. Increasingly, low-cost platforms and custom-made phenotyping hardware are becoming feasible [3, 4].

Here we describe a straightforward, inexpensive method for aboveground imaging of growth chamber or greenhouse-grown plants with a digital camera. The platform uses a commercially available turntable that can be set up in less than a day by nearly any member of the laboratory, and can be designed with one or two cameras for side and/or top views. We routinely use this method for imaging shape, color, and aboveground architecture of tomato plants after biotic stress treatment. In our experience, one person can image approximately 200 plants in 1 day. Resulting images can be analyzed and used with any number of available software packages designed for images in the visual light spectrum, or analyzed with in-house software. We provided tips for optimal lighting, camera settings, tripods, lenses, and plant setup. This is a particularly useful system for researchers new to phenotyping, or for a laboratory that needs to phenotype plants only occasionally but wants high-quality images for downstream analyses.

2 Materials

2.1 Light Studio

1. Commercially available light studio (*see Note 1*).
2. Surge protector power strip.

2.2 Imaging

1. PhotoCapture 360 photography turntable and software by Ortery Technologies (*see Note 2*).
2. Two tables: one large enough to hold the turntable and the other for the computer (*see Note 3*).
3. Computer (*see Note 4*) with USB port access. If your computer has fewer than four USB ports, you will need to buy a multi-port USB hub.
4. Camera(s). Any digital single-lens reflex (DSLR) camera that can be operated with the PhotoCapture software will work (*see Note 5*).
5. Additional camera lenses (*see Note 6*).
6. Additional battery packs and chargers for camera.

7. One interface cable per camera.
8. Tripod for each camera (*see Note 7*).
9. Ethernet cable.
10. USB extension cable.

2.3 Specimen Stage Used in Imaging

1. 100 mm × 100 mm polystyrene square petri dish.
2. Fiducial markers (*see Note 8* and Fig. 1).
3. White foam board 8 × 10 inches (optional; *see Note 9*).
4. Double-sided tape.

3 Methods

3.1 Setting Up the Studio

1. Set a solid worktable in the middle of a room at least 3 m². This is for the PhotoCapture 360 turntable.
2. Place another table in front of the table holding the PhotoCapture 360 and place the laptop on it.
3. Connect the multi-port USB hub to the laptop.
4. Download the PhotoCapture 360 software onto your laptop and set up an account via www.ortery.com. This account will be accessed each time you use the turntable.
5. Position the PhotoCapture 360 photography turntable onto the solid worktable and connect the power supply to a surge protector power supply. Then connect the PhotoCapture 360 to the computer via the supplied cord from Ortery (*see Note 10*).
6. Place a platform cover onto the turntable and position it in the center (*see Note 11*).
7. Set up the light studio per the manufacturer's instructions. Connect all lights to a surge protector power supply.
8. Choose background fabric (*see Note 12*) for your lighting and hang this behind the table. The backdrop stand can be positioned wider than needed to ensure you have an even operating surface.
9. Arrange the light banks them in accordance with Figs. 1 and 2 to start. Then, adjust to your needs.

3.2 Setting Up the Camera

1. Attach the camera lens and fully charged battery to the camera. Attach the quick release plate from the tripod to the camera body. Place camera and quick release plate onto the tripod and ensure it is securely attached.
2. Adjust the tripod to the proper height (*see Note 13*) and distance from the plant sample. Distance will depend on your lens and how close you need to be to your plant to generate the desired image (*see Note 14*).

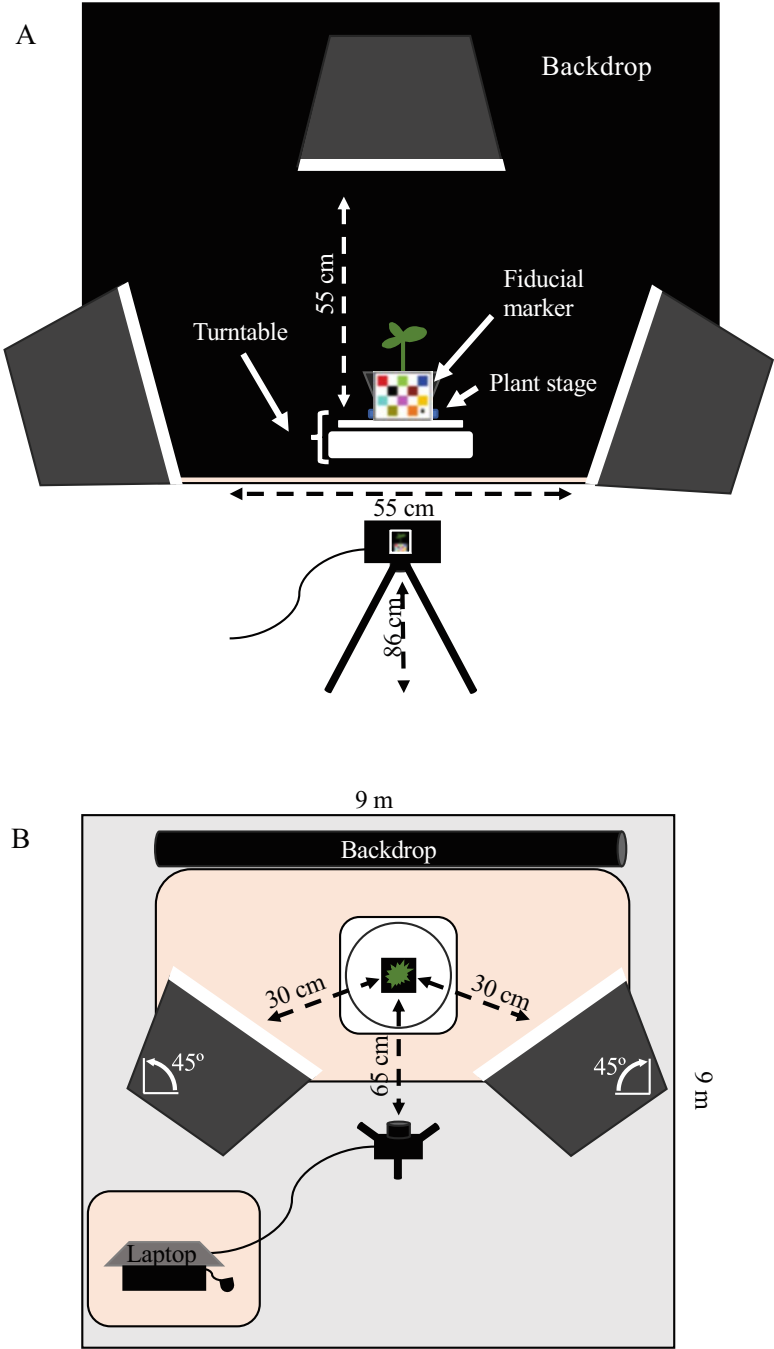


Fig. 1 Platform design with one camera. (a) Side view, (b) top-down view. Distances should be adjusted according to each researcher's needs

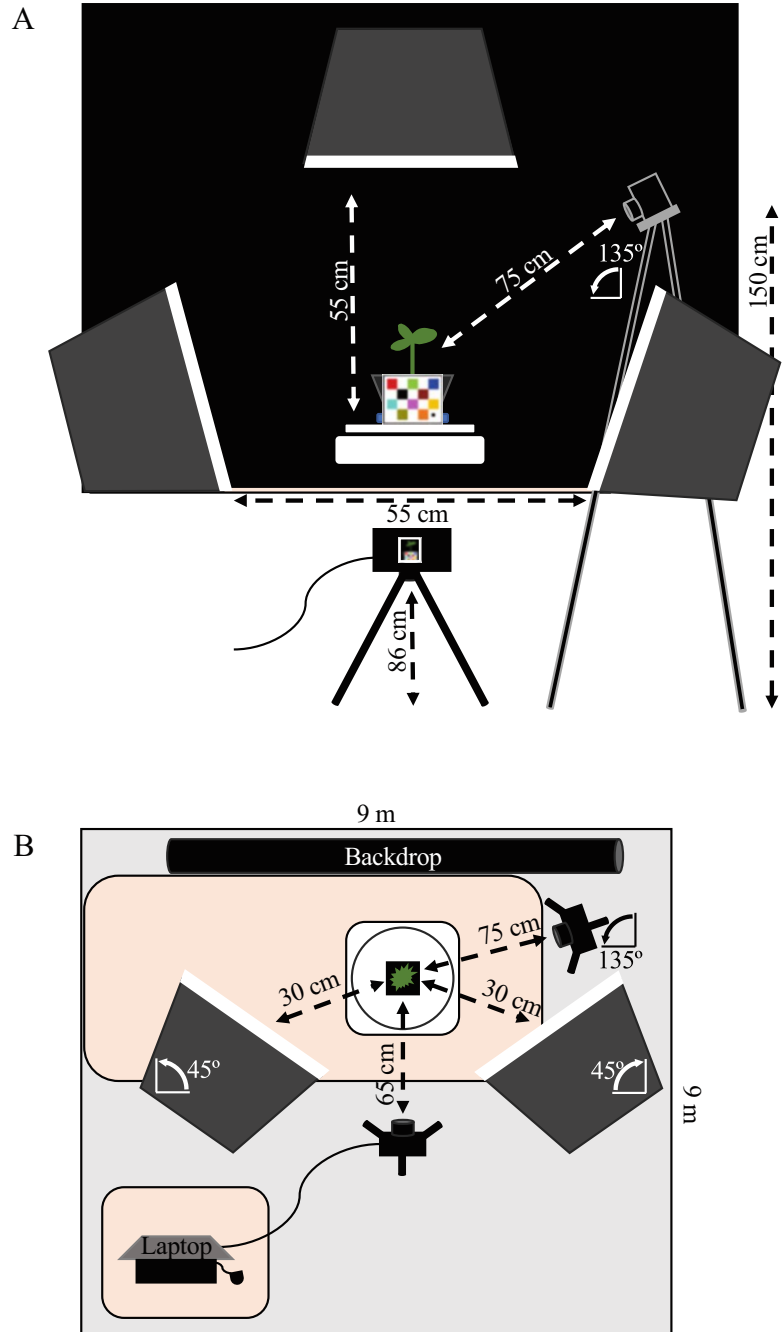


Fig. 2 Platform design with two cameras. (a) Side view, (b) top-down view. Distances should be adjusted according to each researcher's needs

3. Attach the interface cable from the camera to the multi-port USB hub, making sure that the cables do not travel between the camera and the plant sample (*see Note 15*).

4. Place spare battery into the battery charger and connect it the surge protector power supply (*see* **Note 16**).
5. Camera settings: Set the mode dial on the camera to the manual exposure mode (M), ISO speed to 100, shutter speed to 1/30, and the aperture to 11. Set the white balance to color temp and have the lens set to autofocus (AF).
6. If using the two-camera imaging setup, repeat **step 2**, Subheading **3.2**. Lights may need to be adjusted (*see* **Note 17**).

3.3 Setting Up the Plant Stage

1. We use the bottom plate of a 100 mm × 100 mm square petri dish as a stage to hold the plants on the turntable. This ensures plants and fiducial markers are always in the same place (fiducial markers are attached to the sides of the petri dish). We grow plants in cells 9 cm × 9 cm × 6 cm (*see* **Note 18**) and plants of this size easily fit in the petri dish. Place the plant stage in the middle of the turntable.
2. Place fiducial markers on each side of the square petri dish facing out. We affix double-sided tape to the center of each side of the petri dish and stick fiducial markers mounted on foam boards to the tape. If your fiducial markers are printed onto paper, it is helpful to mount them on pieces of white foam board the same size as the markers and then affix the white foam board to each side of the plant stage. The foam board makes the fiducial markers more durable (*see* **Notes 8, 9, and 19**). The reason for placing the marker on each side of the stage is that you will see it in each image you take if you take multiple images around the plant (*see* **Note 20**). Ensure your fiducial marker is oriented the same way on each side of the stage.
3. Mark one side of the ensemble as the front (*see* **Note 21**) and place entire ensemble onto the center of the turntable.

3.4 Imaging

1. Place plant sample into plant stage and align each sample perpendicular to the camera angle (*see* **Note 22**).
2. Open the Ortery Capture software and log in with the username and password you created in **step 4**, Subheading **3.1**. Select Login.
3. Once you are logged in, the software will ask you to set up your workspace. Select Browse and choose your file location. Select OK.
4. After the program has launched, click on the capture button.
5. Under the Capture Settings, select the camera you want to use and fill in your desired camera parameters (*see* **Note 23**).
6. Under the turntable controls selection, choose device, Photo-Capture 360; direction, clockwise; and speed, normal.

7. Set the Animation Settings to spin range, 360, and shooting mode, continuous, and choose the desired number of images around the plant. This will depend on your question (*see Note 20*). Focus camera lens on plant sample. You may adjust focus with the controls on the bottom of the screen.
8. As long as the plant sample is in the center of the turntable and the width of the plant is not too different 360° around the plant, you can select the box that will “Apply current MF settings to all the images in this row.”
9. Once focus is adjusted, select Snap. The camera will now snap the images as it has been programmed. A display of the images being collected will show in the screen and all the thumbnails will be visible below that image.
10. Once all images have been collected, press next and the turntable will return to its starting position.
11. On this screen, you will need to enter the filename for your samples. It is critical to think about this prior to initiating your experiment. The naming system should be uniform. For example, Rep1.Plant1.Day1.FrontCamera, Rep1.Plant2.Day1.FrontCamera (*see Note 24*).
12. To the left of the word Filename is an icon that will add save type files. Click on this and select TIFF and RAW (*see Note 25*).
13. On the right side of the screen is a Browse button so that you can save your file in a specific location. Select save.
14. If using the two-camera system, select the other camera under the Capture Settings-Camera and repeat **steps 5–13**, Subheading [3.4](#).
15. Ideally, it is best to save images directly to a server (*see Note 26*). If you do not do this, upload them after each imaging session.
16. Images can be analyzed using any number of freely available image processing programs (*see Note 27*).

4 Notes

1. We used the Linco Linstore 2000 Watt Photo Studio Lighting Kit with three Color Muslin Backdrop Stand Photography Flora X Fluorescent with four-socket Light Bank and Auto Pop-up Softbox (bought from [Amazon.com](https://www.amazon.com)), but any similar light studio will work. You will need to try different configurations to test what works best for your system.
2. Ortery Technologies (ortery.com) provides a suite of imaging products. The PhotoCapture 360 is a turntable that turns a user-specified number of degrees (from 1° to 360°), stops, and

takes an image. The turntable is controlled by accompanying software that is included in the PhotoCapture 360 cost. The user decides how many images to take, and whether to use one or more cameras.

3. Choose a table for the turntable that is not susceptible to vibration. The operation of the turntable causes a slight vibration that can cause leaf movement if the table is not stable.
4. You can use a desktop or a laptop and either a Mac or a PC as long as it runs Windows 7, 8, or 10. Make sure you look at the Ortery.com webpage before buying the computer and cameras to ensure your models work with the software. We have found laptops are more flexible for confined spaces. We use a Dell Mobile Precision 5530.
5. We use Canon EOS 6D DSLR cameras with a full-frame sensor that allows wider angle images to be taken, but any camera that works with the PhotoCapture 360 software can be used. It is critical to check the ortery.com website to ensure compatibility prior to buying the camera.
6. When considering which lens to buy, think about plant size and size of the room you are using (this will impact distance from sample to camera). We use a Canon EF 50 mm f/1.4 USM lens. If you are working with larger plant sizes, you may need to reconsider the lens. We work with relatively small plants (*see Note 18*) and thus the 50 mm f/1.4 USM lens (fixed focal length) is sufficient. While adjustable zoom lenses would allow a larger selection of plant samples, the cost difference between fixed focal lens and an adjustable zoom lens is large (about US \$1000). The other issue is that the adjustable zoom lens can adjust focal lengths which can cause issues if the tripod location has been moved or the zoom has been manually manipulated.
7. We use Vanguard VEO 265AP Aluminum Travel Tripod with Panning Head from Amazon.com.
8. Fiducial markers are objects used as points of reference and for image registration and orientation. They appear in the same orientation in each image. Depending on the type of fiducial marker, they can also be used for color correction. Many different types of fiducial markers can be used. Depending on your image analysis needs, fiducial markers for images taken with this platform can be as simple as a ruler, or a piece of tape affixed to the stage. We print fiducial markers on heavy paper, and subsequently cut them out and glue them to a white foam board as described in the methods section. You do not need to use the same type of fiducial marker used in Fig. 1. Regardless of what you use as your marker, it is important for downstream analyses that each image has one.

9. The white foam board is used for mounting fiducial markers (if you print them and need to make them more durable) (*see Note 8*).
10. You may need to use a longer cord to connect the computer to the turntable. We found that the turntable did not operate with the USB extension cable, but did connect to the laptop via the multi-port USB hub. We had to try multiple configurations to find the best distance for the laptop from the imaging table.
11. The Ortery turntable comes with four platform covers, two white and two black, and in two different sizes. You can switch them back and forth until you find the cover that works the best for your downstream image analyses. Platform covers are plastic with a hole in the middle. The turntable has a screw in the very center of it. If you line up the hole in the middle of the platform cover with the screw in the center of the turntable, the cover will fit perfectly. Note that the platform cover is not sealed to the turntable. This is beneficial for cleaning as you can treat the cover with ethanol and wipe it up off the platform so that any alcohol does not go into the platform and damage anything.
12. Many commercially available light studios come with different backgrounds. The Linco (*see Note 1*) comes with three backgrounds: white, black, and green. Which background is used depends on the researcher's question. For example, we are interested in biotic stress phenotypes. We use a black background to highlight the color of wilting tomato leaves. This provides sufficient contrast between green healthy leaves and yellow necrotic leaves, but the white background resulted in overexposure of diseased areas of the leaf. If the researcher is interested in leaf lesions, it may be better to use a royal blue background, but each researcher will need to test different backgrounds to determine which is best for their question. If a desired color is not available or the light studio does not come with a background, using a large sheet of cloth of the desired color will work. With all backgrounds, it is critical to be sure that there are not creases in the cloth, as this will make downstream image analysis more challenging.
13. Camera height should be adjusted via the tripod so that the pot is seen directly without observing the surface of the soil. This helps reduce color noise and results in more uniform images. This can be tricky when dealing with tall plants and will need to be adjusted to capture the best possible image, and imaging between plants of different sizes.
14. Tripod distance depends on the lens, size of the plant, and image needed (whole plant, lesion on a leaf, etc.). If you have a fixed focal length lens, the distance of the lens is predefined

and cannot zoom in or out more than is designed. Distance should be the same for each replicate and experimental trial. It is best to set up the camera and test your focal length and then mark the location of the tripod feet with tape on the floor.

15. If using the two-camera system, it may be necessary to use an USB extension cable and run the cords behind or underneath the table to leave the image area unobstructed.
16. This will allow rapid exchange of battery packs during a long photography session without any delay in imaging. We image for 8 h/day and exchange batteries during that time.
17. We have found that one camera that takes images from the side is sufficient for our work. An additional camera taking a top-down image did not provide additional useful information. However, in some species like *Arabidopsis*, imaging top-down may be more beneficial than from the side. The researcher can decide how many cameras and in what positions to place them, depending on the question of interest and images desired. We take images on a daily basis, and measure growth rate over a week. We take four images (all from a side-facing camera) around the plant: one image every 90°. Because plant growth can substantially increase over that period, having four images at different angles around the plant is very helpful for determining leaf order and connections.
18. We image tomato plants grown in 1801 cell packs. This provides 18 single cells in a typical 1020 tray. Each cell is 9 cm × 9 cm and is 6 cm tall. We have imaged plants that vary in height from approximately 9 to 18 cm. Larger plants are possible but adjustments may need to be made to the camera and tripod. Plants smaller than 9 cm height are also fine.
19. Be careful about using liquid glue to mount the markers. You can use a glue stick but we have found that liquid glue will cause the fiducial markers to become distorted.
20. With this software, you can image 360° around your plant sample. An image can be taken each degree if desired. The number of images needed depends on the question—for example, are you interested in 2D or recreating the plant in 3D? If 3D is desired, many more images will be needed. The number of images will also depend on the complexity of your plant sample—i.e., how tall and how many branches and leaves overlap. We take four images (one every 90° around the plant). This is sufficient to determine the correct leaf paths from those that cross in front of one another, but is not sufficient for 3D images. Care should be taken to balance time to image and upload files with the ability to accurately perform downstream data analyses.

21. Make sure the ensemble is positioned with the front mark toward the camera at the beginning of each plant sample. Additionally, ensure that the fiducial marker is perpendicular to the camera at the beginning of imaging each plant sample. This will be important if image registration is needed during image analysis.
22. Prior to the beginning of imaging, each plant sample is marked to indicate which side of the plant will be imaged first. This helps provide consistency in phenotyping.
23. We use the following parameters: focus mode, MF; mode, M; aperture, 11; exposure, 0; white balance, auto; shutter speed; 1/30; and ISO, 100. These settings are a good place to start, but may need to be adjusted depending on your camera, lens, lighting, and plant conditions.
24. Plants should be named with the same naming convention that you will use throughout this experiment and future experiments. This helps greatly with data storage and retrieval. If necessary, plants can be barcoded as well.
25. We routinely save in both file formats. This is helpful for flexibility when sharing images among team members and for image analysis.
26. Regardless of whether using a laptop or desktop, test whether an Ethernet or Internet connection should be used for uploading images to a server (images should be uploaded to a server for storage during imaging or immediately after imaging). We have found that using Ethernet decreases the upload time from 6 min for eight images using Internet connection to 1.5 min for eight images. This decreased time makes a significant difference when imaging hundreds of plants in 1 day.
27. Image processing programs will identify plant pixels within each image, and use these to calculate a range of plant traits. Some programs will also color correct the image, but for these special fiducial markers designed for color correction are necessary if this imaging platform is used. To start phenotyping images, *see* PlantCV, an open-source image analysis software package [9], and also multiple software packages at <https://www.plant-image-analysis.org>. Image processing technology has recently been reviewed in [6]. Image analysis programs will run from manual to semiautomated to automated, but the researchers will need to determine what tools work best for their questions of interest. FIJI/ImageJ also has a number of useful plugins. Regardless of the image analysis program, it is important that the researcher ground truth a set of images, i.e., make a manual measurement and test that against the value obtained with the software package used.

References

1. Furbank RT, Tester M (2011) Phenomics – technologies to relieve the phenotyping bottleneck. *Trends Plant Sci* 16:635–644
2. Tardieu F, Cabrera-Bosquet L, Pridmore T et al (2017) Plant phenomics, from sensors to knowledge. *Curr Biol* 27:R770–R783
3. Araus JL, Kefauver SC (2018) Breeding to adapt agriculture to climate change: affordable phenotyping solutions. *Curr Opin Plant Biol* 45: 237–247
4. Fahlgren N, Gehan MA, Baxter I (2015) Lights, camera, action: high-throughput plant phenotyping is ready for a close-up. *Curr Opin Plant Biol* 24:93–99
5. Simko I, Jimenez-Berni JA, Sirault XRR (2017) Phenomic approaches and tools for phytopathologists. *Phytopathology* 107:6–17
6. Perez-Sanz F, Navarro PJ, Egea-Cortines M (2017) Plant phenomics: an overview of image acquisition technologies and image data analysis algorithms. *GigaScience* 6:1–18
7. Atkinson JA, Pound MP, Bennett MJ et al (2019) Uncovering the hidden half of plants using new advances in root phenotyping. *Curr Opin Biotechnol* 55:1–8
8. Paez-Garcia A, Motes C, Scheible WR et al (2015) Root traits and phenotyping strategies for plant improvement. *Plan Theory* 4:334–355
9. Gehan MA, Fahlgren N, Abbasi A et al (2017) PlantCV v2: image analysis software for high-throughput plant phenotyping. *PeerJ* 5:e4088



Wireless Fixed Camera Network for Greenhouse-Based Plant Phenotyping

Nadia Shakoor and Todd C. Mockler

Abstract

An indoor wireless fixed camera network was developed for an efficient, cost-effective method of extracting informative plant phenotypes in a controlled greenhouse environment. Deployed at the Donald Danforth Plant Science Center (DDPSC), this fixed camera platform implements rapid and automated plant phenotyping. The platform uses low-cost Raspberry Pi computers and digital cameras to monitor aboveground morphological and developmental plant phenotypes. The Raspberry Pi is a readily programmable, credit card-sized computer board with remote accessibility. A standard camera module connects to the Raspberry Pi computer board and generates eight-megapixel resolution images. With a fixed array, or “bramble,” of Raspberry Pi computer boards and camera modules placed strategically in a greenhouse, we can capture automated, high-resolution images for 3D reconstructions of individual plants on timescales ranging from minutes to hours, capturing temporal changes in plant phenotypes.

Key words Greenhouse, Phenotyping, Raspberry Pi, 3D reconstruction, Image analysis, Imaging

1 Introduction

A wireless fixed camera array was developed and deployed at the Donald Danforth Plant Science Center (DDPSC) to monitor aboveground morphological and developmental plant phenotypes. One hundred eighty Raspberry Pis (Pis) and connected camera modules (Fig. 1) are positioned on an overhead scaffold in a grid formation. The camera array, or “bramble,” is powered by blocks of electrical power strips with USB adaptors and cables, and network access is provided via WiFi dongles attached to each Pi. The Pis are affixed to the scaffold 3 meters above the ground with flexible tripods (Fig. 2). Each camera is angled to provide overlapping fields of view to ensure high-quality 3D reconstructions.

3D reconstructions and image analyses are carried out in VisualSFM [1] and CloudCompare [2], respectively (Fig. 3). Using the connected component analysis feature in

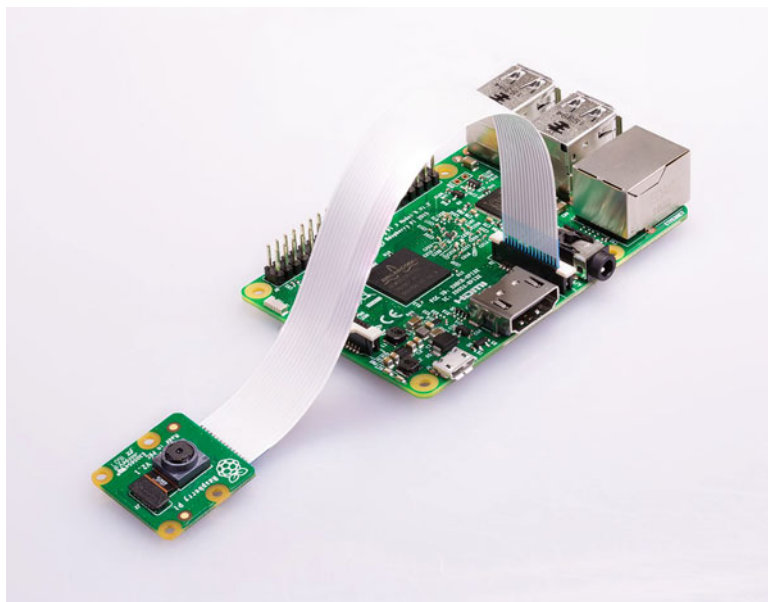


Fig. 1 Raspberry Pi 3 with an attached 8 MP camera module (www.raspberrypi.com)



Fig. 2 Raspberry Pi camera setup (case, WiPi dongle, power cord) from two views attached to scaffold with a flexible tripod

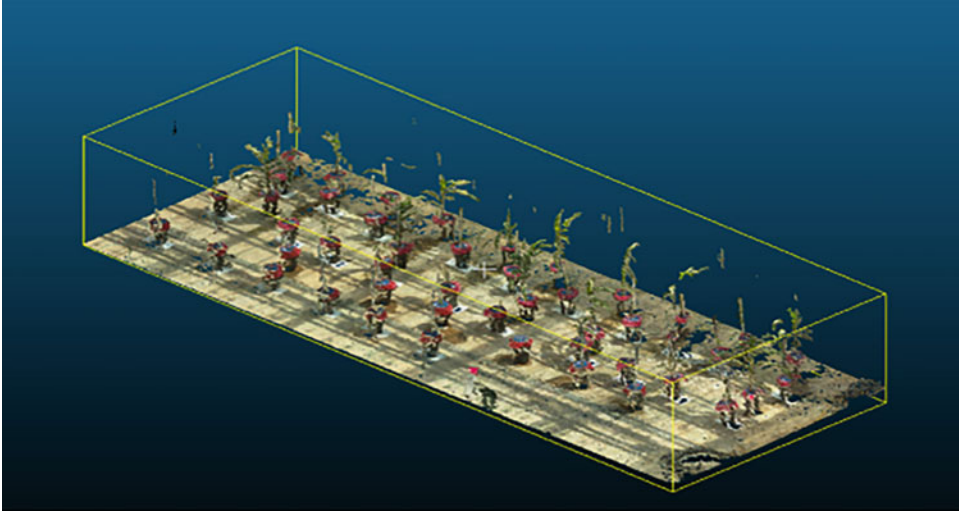


Fig. 3 VisualSFM 3D reconstruction and CloudCompare visualization of the greenhouse point cloud

CloudCompare, we can bound and separate each plant in the greenhouse and accurately calculate measures of biomass accumulation, plant height, leaf area, leaf angle, and growth rate. Comparing ground-truth and image-derived measurements of plant heights, we attain 1–2 cm accuracy in distance measurements with R-squared values of greater than 0.9 (Fig. 4a–c).

2 Materials

- Raspberry Pi 3 computers (*see Note 1*).
- Raspberry Pi 8MP camera modules (*see Note 1*).
- Raspberry Pi camera cases.
- Flexible tripods.
- WiPi WiFi dongles.
- MicroSD cards (minimum 8 GB storage).
- USB power cords.
- Electrical power strips.
- Wireless access point(s)

3 Methods

3.1 Hardware Setup

Raspberry Pis and connected camera modules are enclosed in a water-resistant plastic case and are affixed to an overhead scaffold in an evenly spaced grid formation. WiPi WiFi dongles are used to connect the Pis to the WiFi local area network. Raspberry Pi

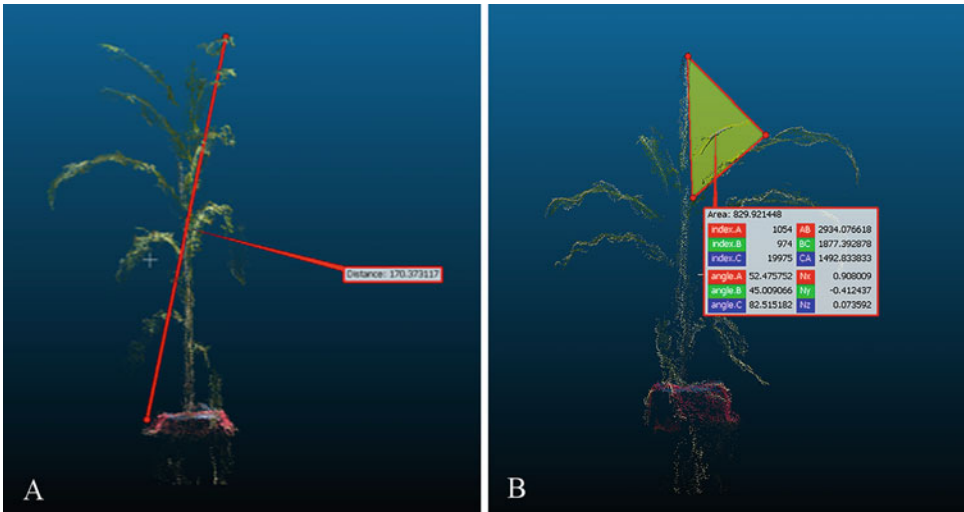


Fig. 4 CloudCompare feature analysis of a single plant with derived plant height (a), leaf angle (b), and leaf lengths (c)

3 computers have internal WiFi; however, we found that using an external WiPi dongle improves signal strength and stability. The camera modules are angled using flexible tripods to ensure 20–30% overlap in the fields of view between adjacent cameras.

3.2 Software and Management

The Raspberry Pi camera array is online at all times and was created using Raspbian, a Debian-based operating system for Raspberry Pi software (www.raspbian.org). Ganglia [3] is used to obtain real-time reporting data from the bramble, including a visualization of the Raspberry Pis that are currently online and transmitting data (Fig. 5). We also recommend spot checking the images downloaded at a given time point or accessing each Pi remotely to verify connectivity and transmission. Independent of network access, image capture is initiated on each Pi camera system via a cron job that is set up on the Raspberry Pi device. Images are stored on a local SD card and then imported to local storage, or a compute infrastructure from each Raspberry Pi using a cron job set up on a rsync server. The Raspberry Pi bramble is managed from the server using a deployment configuration engine called Ansible [4]. Each Pi’s configuration requires a unique hostname and IP address. Additional requirements include WiFi access either via a WiPi dongle or use of Raspberry Pi 3 camera, a set time zone, and an OpenSSH server. A microSD card with a minimum of 8 GB storage is recommended for each Pi system to accommodate system specifications and local image storage.

Wireless power management is a standard feature on the Raspberry Pi that disables the WiFi dongle after periods of inactivity. This setting can be turned off using the code “sudo nano /etc/



Fig. 5 Ganglia interface showing CPU load of each Pi in the bramble. CPU load for four camera systems are individually highlighted in bottom panels

network/interfaces” and entering “wireless-power off” in the wlan0 section. OpenSSH server is also installed on each Pi using the command “sudo apt-get install openssh-server.” OpenSSH allows for remote access and transfer capabilities. We also recommend setting the time zone for each Pi using the command “sudo dpkg-reconfigure tzdata.” A local image directory is also created using the command “mkdir /home/pi/Images” which creates a directory named “Images” on the Pi.

3.3 Raspberry Pi Bramble Setup

We built an array of 180 Raspberry Pis, which provides sufficient coverage and 20–30% image overlaps for a 1400 sq ft/130 m² greenhouse (25' × 56'/7.6 m × 17.1 m). The Pis are affixed in a grid arrangement to a scaffold (15' × 45'/4.6 m × 13.7 m) 3 m above the ground.

3.4 Calculation of Distance Units

A GPS reference point from the corner of the greenhouse was identified using Google Maps/Earth Mercator projection. QGIS [5] was used to assign GPS points to the Raspberry Pis in the bramble given the reference GPS location. In VisualSFM, GPS points were added to the 3D reconstruction images, which translates point cloud units into centimeters for visualization and analysis in CloudCompare (Fig. 6).

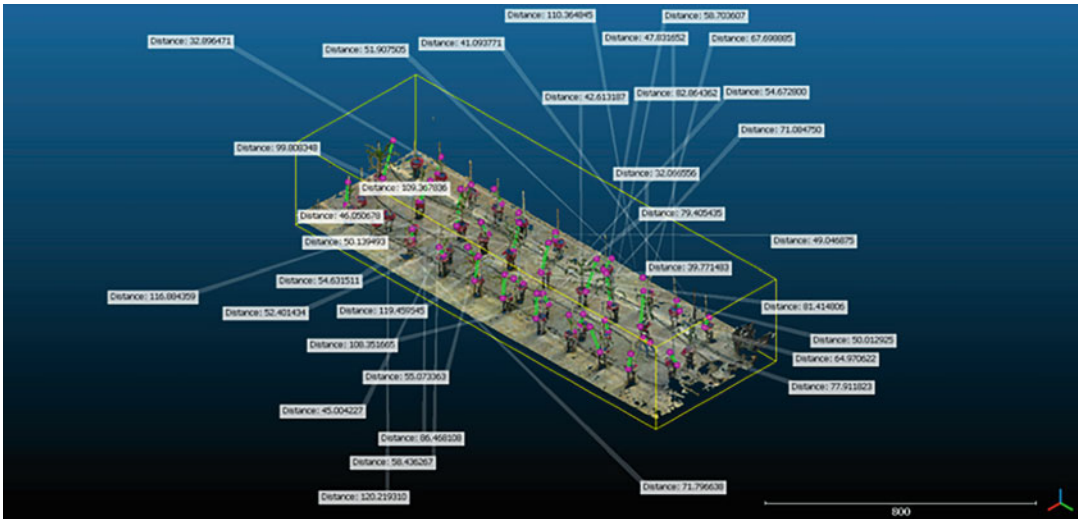


Fig. 6 CloudCompare distance measurements in centimeters

3.5 3D Reconstruction and Analysis in VisualSFM and CloudCompare

VisualSFM is used to reconstruct the interior of the greenhouse, including all plants and pots in the space. Using connected component analysis in CloudCompare, the 3D reconstruction is separated into individual point clouds. Thresholding is used to determine the minimum number of points per component, which subsequently selects the smallest set of points that can be considered a “connected component” (Fig. 7a, b). Complete objects with multiple point clouds (e.g., several leaves on one plant) require manual merging.

4 Notes

1. It is recommended that the newest, highest-resolution Raspberry Pi computers and camera modules available be used when building a camera network.
2. The most significant challenge with the wireless fixed camera system is maintaining a sustained connection to each Raspberry Pi for the amount of time it takes to transfer a file. Wireless interference is the typical cause of this issue. There are four wireless access points (WAPs), each connected to 45 Pis, stationed in and around the greenhouse that cause interference; the 180 Raspberry Pis themselves also cause interference.
3. We implemented several methods to reduce the impact of wireless interference. Four WAPs and four SSIDs were used exclusively for the bramble, preventing Raspberry Pi network dropout. WiPi dongles also helped with establishing connectivity and maintaining connection stability. We also decreased the transmit power from each Raspberry Pi until it

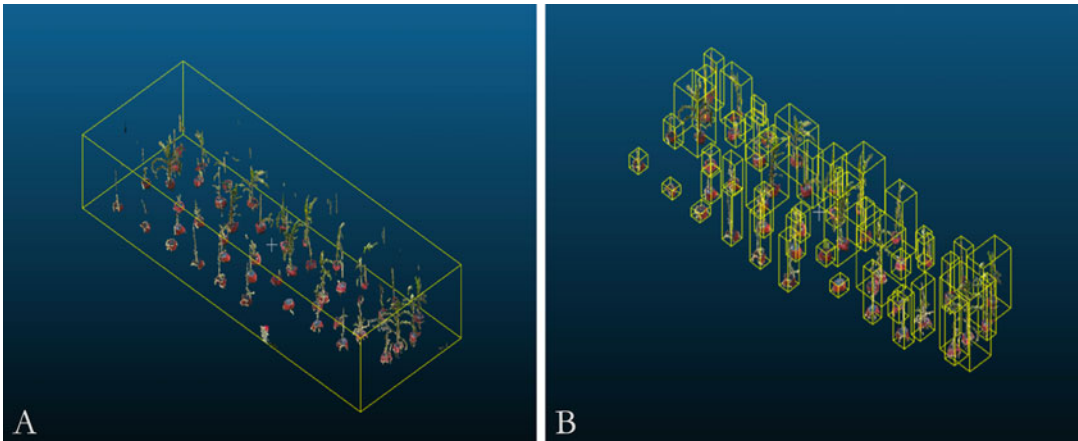


Fig. 7 CloudCompare connected component analysis pipeline with thresholding (a) and connected component extraction (b)

was time to transfer the files. The transmit power was increased for the duration of the transfer time.

4. Hardwiring: If feasible, each Raspberry Pi can be hardwired to a switch that is directly connected to the network.
5. Partial hardwiring: We recommend partially hardwiring the Raspberry 5. Pis, six at a time to a wireless bridge, bringing the number of devices connected to each SSID/WAP to eight Pis per SSID/WAP instead of 45 per SSID/WAP.
6. WiFi dongles with directional antennas: Dongles with directional antennas may also aid in reducing wireless interference.
7. The quality of the Pi-generated input images in VisualSFM determines the quality of the 3D reconstructions. Poor image quality in the greenhouse is primarily due to environmental variables, including variations in sunlight passing through the greenhouse roof and resulting in under- and oversaturated images. Cloudy days or shadows from clouds or overhead lighting fixtures can also dramatically reduce image quality and the resolution of the subsequent 3D reconstruction.
8. Several image processing methods can be applied to the Pi-generated images before the 3D reconstruction step. These include thresholding, machine learning-based classification methods, and image enhancements [6–8]. Addition of color markers or other distinct objects in the greenhouse can also aid in the reconstruction process. We have found that VisualSFM produces a better reconstruction output when there are unique features captured in the greenhouse during image acquisition. Additionally, we find that increasing the rate of image acquisition allows for the removal of sub-par 3D reconstructions without compromising the phenotyping experiment.

Acknowledgments

This protocol reflects the input of a larger project team including César Lizárraga, Brandon Patrick, Phil Ozersky, Stuart Marshall, Bradley Flynn, Avishek Datta, and Darren O'Brien. The information presented here is based upon work partially supported by the National Science Foundation under Award Number IIA-1355406 and the US Department of Energy Advanced Research Projects Agency-Energy (ARPA-E) under Cooperative Agreement Number DE-AR0000594. The views and opinions of authors expressed herein do not necessarily state or reflect those of the US government or any agency thereof.

References

1. Wu C (2011) VisualSFM: a visual structure from motion system
2. CloudCompare (version 2.1) (2019) Retrieved from <https://www.cloudcompare.org/>
3. Massie ML, Chun BN, Culler DE (2004) The ganglia distributed monitoring system: design, implementation, and experience. *Parallel Comput* 30(7):817–840
4. Hochstein L, Moser R (2017) Ansible: up and running: automating configuration management and deployment the easy way. O'Reilly Media, Inc
5. QGIS Development Team. (2016) QGIS geographic information system. Open Source Geospatial Foundation Project. Retrieved from <https://www.osgeo.org/>
6. Yu K, Kirchgessner N, Grieder C et al (2017) An image analysis pipeline for automated classification of imaging light conditions and for quantification of wheat canopy cover time series in field phenotyping. *Plant Methods* 13(1):15
7. Liebisch F, Kirchgessner N, Schneider D et al (2015) Remote, aerial phenotyping of maize traits with a mobile multi-sensor approach. *Plant Methods* 11(1):9
8. Hartmann A, Czauderna T, Hoffmann R et al (2011) HTPheno: an image analysis pipeline for high-throughput plant phenotyping. *BMC Bioinform* 12(1):148



Experimental Design for Controlled Environment High-Throughput Plant Phenotyping

Jennifer L. Clarke, Yumou Qiu, and James C. Schnable

Abstract

It is essential that the scientific community develop and deploy accurate and high-throughput techniques to capture factors that influence plant phenotypes if we are to meet the projected demands for food and energy. In recognition of this fact, multiple research institutions have invested in automated high-throughput plant phenotyping (HTPP) systems designed for use in controlled environments. These systems can generate large amounts of data in relatively short periods of time, potentially allowing researchers to gain insights about phenotypic responses to environmental, biological, and management factors. Reliable inferences about these factors depends on the use of proper experimental design when planning phenotypic studies in order to avoid issues such as lack of power and confounding. In this chapter, the topic of experimental design will be discussed, from basic principles to examples specific to controlled environment plant phenotyping. Examples will be provided based on the package *agricolae* in the R statistical language.

Key words Design of experiments, *Agricolae*, Genomes2Fields, G2F

1 Introduction

According to [1], a plant phenotype is “a quantitative description of the plants anatomical, ontogenetical, physiological and biochemical properties.” Scientists and farmers alike are interested in phenotypes as quantitative assessments of plant performance, where performance may be defined as overall yield or stress tolerance or response to data-driven management. Scientists are also interested in phenotypes as reflections of how plant life is organized across scales, from molecular to ecological [2]. Plant responses such as growth are best measured over time, leading to interest in nondestructive approaches to phenotyping. Modern high-throughput plant phenotyping (HTPP) systems can measure multiple

Dedication: I would like to dedicate this article to Professor Argelia Lorence for her commitment to inclusion in science and academia.

phenotypes over time, nondestructively, in either controlled or field conditions. These systems can generate large amounts of data relatively quickly, which can lead researchers to assume that such data contains plenty of scientifically valuable information. However, reliable and reproducible inferences from such data depend on the use of proper experimental design in order to avoid issues such as lack of power and confounding. The expense and high demand associated with controlled environment HTPP make proper design a critical consideration for any study.

Although HTPP is relatively new, experimental design has been around for decades; see Box, Hunter, and Hunter [3], Gómez and Gómez [4], and Montgomery [5]. This chapter will provide a brief overview of experimental design including key principles, optimality and efficiency, and some designs common to agricultural research. Included is an example of design used for a HTPP study associated with the Genomes to Fields (G2F) Initiative [6] (<https://www.genomes2fields.org/home>). Code to implement various designs and examples is provided based on the R statistical language [7] and the agricolae package (version 1.3-1) [8] available from the Comprehensive R Archive Network (CRAN).

2 Overview of Experimental Design

Experimental design is a way to plan experiments in advance so that the results are as objective and valid as possible. Ideally, an experimental design should achieve the following: (1) describe how participants are allocated to experimental groups, (2) minimize or eliminate confounding variables, (3) permit inferences about the relationships between independent and dependent variables, and (4) reduce variability which facilitates the finding of differences in treatment outcomes. A design rests on the following principles:

- Randomization: Eliminates bias from the results. Assign individuals to treatments using a random method.
- Replication: Experiment must be replicable by other researchers. Use statistics like standard error of the sample mean or confidence intervals. Blocking: controlling sources of variation.

When discussing the design of experiments (DoE), statisticians often focus on two aspects, namely, optimality and efficiency [9]. An optimal design allows parameters to be estimated without bias and with minimum variance (note: bias is related to inaccuracy and variance is related to imprecision). The measurement of optimality depends on the statistical model of the data and the specific optimality criterion. The efficiency of a model is usually stated as relative to the best (i.e., least variable) model, so a ratio of the minimal possible variance to the actual variance. For the most efficient model, this ratio will equal one, and this ratio decreases with decreasing efficiency.

2.1 Commonly Used Designs

This section will briefly cover several common designs of increasing complexity through examples, with associated R code. For computational implementation of any of the code below, the first step is to download and load the *agricolae* package within R. The package can be downloaded from <https://cran.r-project.org/web/packages/agricolae/index.html>.

```
install.packages("agricolae")
library(agricolae)
designs <- apropos("design")
print(designs[substr(designs,1,6)=="design"], row.names=FALSE)
```

2.2 Completely Randomized Design (*design.crd*)

A completely randomized design (CRD) has treatments assigned at random. In other words, every experimental unit has the same probability of receiving a given treatment. As an example, the sweetpotato data in *agricolae* correspond to a CRD with 12 plots, 50 sweet potato plants per plot, 4 treatments (CC (Spcsv) = Sweetpotato chlorotic dwarf, FF (Spfmv) = Feathery mottle, FC (Spfmv y Spcsv) = Viral complex and OO (witness) healthy plants), and 3 replicates. Note that this example has one treatment per plot. The relevant command for CRD is *design.crd*.

```
data(sweetpotato)
str(sweetpotato)
trt<-levels(sweetpotato$virus)
r<-c(rep(3, 4))
outdesign1<-design.crd(trt,r,serie=0,seed=2020)
book1<-outdesign1$book
head(book1)
write.table(book1, "crd.txt", row.names=FALSE, sep="\t")
```

The first part of the resulting design (output from the above) will be as follows:

plots r trt

```
1 1 1 fc
2 2 1 ff
3 3 2 ff
4 4 1 oo
5 5 2 oo
6 6 1 cc
```

Here is what the design looks like (*see* Table 1).

The full design will be output and saved to the file *crd.txt* in the working directory.

Table 1
Completely randomized design (design.crd)

1	2	3	4
fc	ff	ff	oo
5	6	7	8
oo	cc	oo	cc
9	10	11	12
fc	cc	fc	ff

**2.3 Randomized
Complete Block Design
(design.rcbd)**

In a randomized complete block design (RCBD), the researcher divides experimental subjects into homogeneous blocks and assigns treatments randomly *within* each block. Under this model, it is assumed that variability within blocks is larger than variability between blocks. This design avoids potential confounding by the blocking variable. All treatments appear exactly once within each block. The grass data correspond to a RCBD with 12 lawns each with 4 subplots, and 4 treatments (type of grasses) with, as expected, one treatment per subplot. The relevant command for RCBD is `design.rcbd`.

```
data(grass)
str(grass)
trt<-levels(grass$trt)
r<-12
outdesign<-design.rcbd(trt,r,serie=2,seed=2020)
book2<-outdesign$book
head(book2)
write.table(book2, "rcbd.txt", row.names=FALSE, sep="\t")
```

The first part of the resulting design (output from the above) will be as follows:

plots block trt

1	101	1	t2
2	102	1	t4
3	103	1	t3
4	104	1	t1
5	201	2	t3
6	202	2	t2

Here is what the design looks like (*see* Table 2).

The full design will be output and saved to the file `rcbd.txt` in the working directory.

Table 2
Randomized complete block design (design.rcbd)

101	102	201	202	301	302	401	402
t2	t4	t3	t2	t3	t1	t4	t2
103	104	203	204	303	304	403	404
t3	t1	t1	t4	t4	t2	t1	t3
501	502	601	602	701	702	801	802
t2	t1	t4	t3	t2	t3	t1	t3
503	504	603	604	703	704	803	804
t4	t3	t2	t1	t1	t4	t4	t2
901	902	1001	1002	1101	1102	1201	1202
t4	t2	t1	t3	t3	t4	t1	t2
903	904	1003	1004	1103	1104	1203	1204
t1	t3	t2	t4	t1	t2	t3	t4

2.4 Balanced Incomplete Block Design (design.bib)

A balanced incomplete block design (BIB) may be used when each pair of treatments occur together λ times and not all treatments fit into each block. With a treatments, b blocks, r replicates, and k treatments per block, the total number of observations is $kb = ar = N$. Under this design, each treatment occurs in r blocks. To have balance, each treatment has equal probability of occurring with every other treatment. With $(k - 1)$ other treatments in a block and $(a - 1)$ other treatments, $\lambda = r(k - 1)/(a - 1)$.

As an example of a BIB design, revisit the grass data with 12 lawns each with 4 subplots, and 4 treatments (type of grasses; $a = 4$), one treatment per subplot. What if $k = 2$ and $b = 6$? In this case, $r = 3$ and $\lambda = 1$. The relevant command for BIB is design.bib.

```
str(design.bib) function (trt, k, r = NULL, serie = 2, seed =
0, kinds = "Super-Duper", maxRep = 20, randomization = TRUE)
trt< levels(grass$trt) k<-2
```

```
outdesign<-design.bib (trt,k,r=3,serie=2,seed=2020)
book3<-outdesign$book
```

```
write.table(book3, "bib.txt", row.names=FALSE, sep="\t")
```

Below is some information about the resulting design:

Parameters BIB	
=====	
Lambda	: 1

(continued)

Table 3
Balanced incomplete block design (design.bib)

101	102	201	202	301	302
t1	t2	t4	t2	t2	t3
401	402	501	502	601	602
t4	t3	t3	t1	t4	t1

treatmeans	: 4
Block size	: 2
Blocks	: 6
Replication	: 3
Efficiency factor 0.6666667	

Note that the output above provides the *efficiency* factor as mentioned in Subheading 2. The efficiency of this design is less than one as this is an incomplete design, i.e., we are trading away some efficiency in exchange for a smaller design. This can be important when observations are expensive. Here is what the design looks like (*see* Table 3).

The full design will be output and saved to the file bib.txt in the working directory.

2.5 Other Incomplete Designs

The better designs (in terms of efficiency) have every pair of treatments occurring the same, or nearly the same, number of times in the row blocks and column blocks. Some commonly used designs are briefly listed here.

- Youden square (design.youden) is a Latin square with one row (col) deleted so each treatment occurs in each row (col).
- Partially balanced incomplete block design (PBIB) (design.alpha) doesn't require each pair of treatments to occur λ times. Instead, each pair of treatments in a defined associate class i appear together λ_i times. For general alpha designs, $k < b$, $a < b \times k$.
- Cyclic design (design.cyclic) may be relevant in situations where $r = mk$ and $b = ma$. This is a very large class and includes some BIB and PBIB designs.
- Lattice design (design.lattice) includes square ($a = k^2$), cubic ($a = k^3$), and rectangular ($a = k(k + 1)$) lattices. For an example of a lattice design, *see* He et al. [10].

An example of the above designs is the partially balanced incomplete block design (design.alpha). Consider 30 treatments/

Table 4
Partially balanced incomplete block design (design.alpha)

101	102	103	104	105	106	107	108	109	110	111	112
g16	g20	g7	g24	g17	g11	g10	g28	g18	g29	g12	g14
113	114	115	116	117	118	119	120	121	122	123	124
g4	g6	g9	g2	g27	g15	g1	g26	g30	g13	g19	g5
125	126	127	128	129	130						
g8	g23	g21	g25	g3	g22						
						201	202	203	204	205	206
						g10	g13	g12	g19	g18	g21
207	208	209	210	211	212	213	214	215	216	217	218
g6	g22	g17	g7	g15	g1	g3	g24	g27	g14	g9	g28
219	220	221	222	223	224	225	226	227	228	229	230
g11	g4	g29	g20	g2	g25	g8	g5	g26	g23	g30	g16

genotypes, 2 repetitions, and a block size equal to 3. (In other words, $a = 30$, $k = 3$, $r = 2$.)

```
Genotype<-paste("geno",1:30,sep="")
r<-2
k<-3
plan<-design.alpha(trt=Genotype,k=k,r=r,serie=2,seed=2020)
book4<-plan$book
write.table(book4, "pbib.txt", row.names=FALSE, sep="\t")
```

Below is some information about the resulting design:

Treatmeans: 30

Block size: 3

Blocks: 10

Replication: 2

Efficiency factor (E) 0.6170213

Here is what the design looks like (*see* Table 4).

The full design will be output and saved to the file pbib.txt in the working directory.

3 Analysis of DoE: PBIB

Analysis of the designs presented here is usually through analysis of variance (ANOVA) tables and restricted maximum likelihood (REML). By these methods, one can examine effects due to

treatment as well as other factors. In cases with lots of treatments and tests of significance, multiple comparison corrections are recommended; *see* refs. [11] and [12] for more information.

Assume the experimental study investigating yield as a response to genotype and the associated experimental design as described before. The code for the analysis of that design may be as follows:

```
yield <- c(5,2,7,6,4,9,7,6,7,9,6,2,1,1,3,2,4,6,7,9,8,7,6,4,3,2,2,1,1,2,1,1,2,
4,5,6,7,8,6,5,4,3,1,1,2,5,4,2,7,6,6,5,6,4,5,7,6,5,5,4) data<-data.
frame(plan$book,yield)
modelPBIB<-with(data,PBIB.test(block,Genotype,replication,
yield,k=3,console=TRUE))
```

Requests for various outputs from the model may include the following:

Parameter Estimates (e.g., `head(modelPBIB $ means)`)

Fit Statistics (e.g., `head(modelPBIB $ comparison)`)

Analysis of Variance Table head (`modelPBIB$comparison`) (e.g., `summary(modelPBIB)`)

4 G2F Greenhouse Study

The goal of this study, conducted in 2017 in the Greenhouse Innovation Complex at the University of Nebraska-Lincoln, was to assess the heritability and spatial/temporal effects in a high-throughput automated greenhouse environment on maize biomass using the Genomes2Fields (G2F) diversity panel [6]. Thirty-two maize inbreds were used, including B73 and 31 lines grown and phenotyped under a variety of field conditions in 2014 and 2015. The associated field data from 2014 is available from <https://doi.org/10.7946/P2201Q>, and the RBG images from the greenhouse study are available from <http://plantvision.unl.edu/>. Imaging started 8 days after planting, with watering to target a weight of 540 g, and continued until day 39 after planting.

The area available in the greenhouse for this study contained 96 pots arranged in 8 rows and 12 columns. With 32 genotypes, the question arose regarding the optimal design, i.e., the choices of b and r . As each experimental unit would be relatively expensive and limited resources were available, the choice was made to use an incomplete block design (`design.alpha`) in which tables of blocks are rectangular, col by row , and $row < col$. The number of treatments, i.e., genotypes, was $b \times k = 32$ and the number of experimental units was $r \times b \times k = 96$. This type of incomplete block design is known as an alpha lattice design [10]. Alpha lattice designs are *resolvable*, i.e., the incomplete blocks group together into super-blocks that are complete. The results can be analyzed with ANOVA (and restricted maximum likelihood) or functional modeling approaches [13–15].

The following code will generate an alpha lattice design for this study:

```
trt<-c(paste("t",1:32,sep=""))
design.alpha(trt=trt,k=4,r=3,serie=1,seed=2002)
plan<-outdesign$book
write.table(plan, "alphalattice.txt", row.names=FALSE, sep="\t")
```

Below is some information about the resulting design:

Alpha Design (0,1) – Serie III
Parameters Alpha Design
=====
Treatments: 32
Block size: 4
Blocks: 8
Replication: 3
Efficiency factor
(E) 0.746988

Here is what the design looks like (*see* Table 5).

Table 5
Alpha lattice design (design.alpha)

11	12	13	14	11	12	13	14	11	12	13	14
g28	g16	g4	g2	g1	g4	g5	g18	g13	g15	g31	g29
21	22	23	24	21	22	23	24	31	32	33	34
g13	g21	g25	g30	g3	g27	g17	g29	g10	g5	g9	g28
31	32	33	34	31	32	33	34	41	42	43	44
g10	g31	g20	g8	g6	g30	g7	g31	g18	g30	g14	g20
41	42	43	44	41	42	43	44	41	42	43	44
g27	g7	g9	g19	g10	g11	g23	g19	g2	g1	g27	g11
51	52	53	54	51	52	53	54	51	52	53	54
g32	g6	g1	g12	g12	g16	g20	g21	g8	g4	g26	g7
61	62	63	64	61	62	63	64	61	62	63	64
g18	g3	g15	g23	g26	g32	g25	g2	g24	g12	g3	g25
71	72	73	74	71	72	73	74	71	72	73	74
g11	g17	g24	g22	g9	g13	g22	g8	g17	g6	g23	g21
81	82	83	84	81	82	83	84	81	82	83	84
g29	g5	g26	g14	g14	g15	g28	g24	g32	g19	g22	g16

The full design will be output and saved to the file `alphaLattice.txt` in the working directory.

For the analysis of this design, effects of interest include block and genotype effects on a given trait. The statistical model can be expressed as

$$y_{ij,t} = \mu_t + \alpha_{i,t} + \gamma_{\nu(i,j),t} + \varepsilon_{ij,t},$$

where i = block, $\nu(i,j)$ = plant j in block i , t = time. This model can be used to examine effects due to treatment (genotype).

Assume that biomass is the outcome of interest. Then some code for the data analysis is as follows:

```
biomass <- c(biomass)
data <- data.frame(outdesign$book, biomass)
modelPBIB <- with(data, PBIB.test(block, trt, replication, biomass,
k=4, console=TRUE))
plot(modelPBIB, las=2)
```

The resulting plot looks something like Fig. 1 depending on the values of biomass.

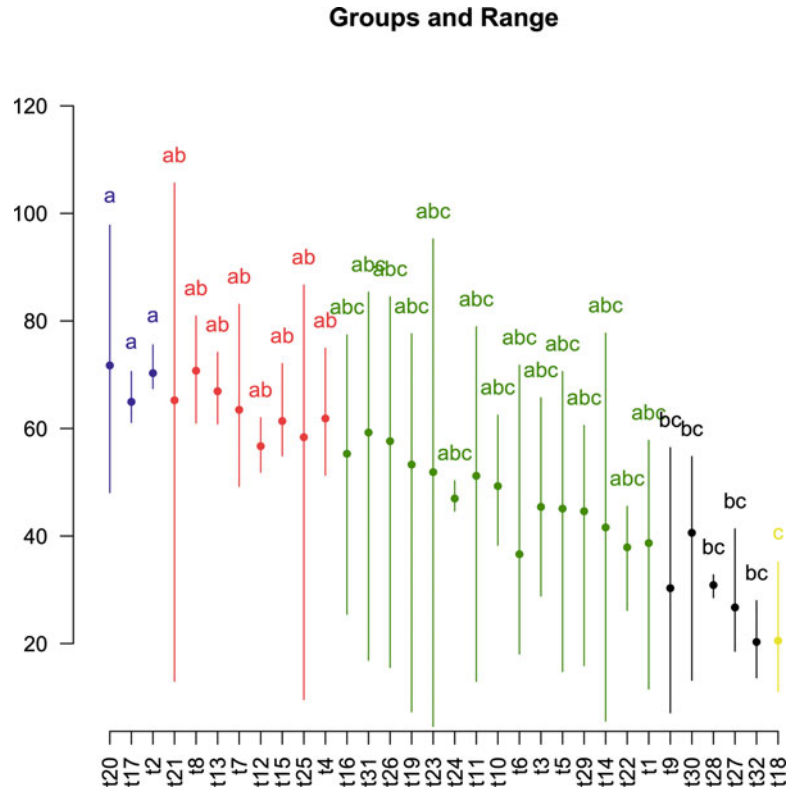


Fig. 1 Analysis of G2F alpha lattice design

In Fig. 1 the genotypes are noted along the x -axis and biomass is the y -axis. Groups of genotypes with significant differences in biomass are shown in different colors with different group labels.

5 Conclusions

Experimental design is a critical part of any research study. Use of a proper design provides some level of confidence that analyses of the resulting data will be sufficient to answer the scientific hypotheses. The designs presented in this chapter have a long history in agriculture and apply in the modern setting of HTPP. Although greenhouses provide a higher level of environmental control than most field conditions, there remains variability within the greenhouse microclimate that can impact the results of a study. The use of statistical design can help separate the effects of treatments/conditions/genotypes from other potentially confounding factors. This improves reproducibility of any findings.

In addition to the agricolae package, the R language has several other packages that contain tools for experimental design (such as desplot [16] and data from agridat [17]). A reliable source for information about such packages is the CRAN Task View on Design of Experiments and Analysis of Experimental Data (<https://CRAN.R-project.org/view=ExperimentalDesign>). This page contains a special section on designs for agricultural and plant breeding experiments as well as a list of key references.

References

1. Walter A, Liebisch F, Hund A (2015) Plant phenotyping: from bean weighing to image analysis. *Plant Methods* 11(1):14. <https://doi.org/10.1186/s13007-015-0056-8>
2. National Science Foundation (2017) Rules of life outlines path to predicting phenotype. NSF News Release
3. Box G, Hunter J, Hunter W (2005) *Statistics for experimenters: design, innovation, and discovery*, 2nd edn. Wiley-Interscience
4. Gómez K, Gómez A (1984) *Statistical procedures for agricultural research*, 2nd edn. Wiley
5. Montgomery D (2013) *Design and analysis of experiments*, 8th edn. Wiley
6. Gage J, Diego Jarquin H et al (2017) The effect of artificial selection on phenotypic plasticity in maize. *Nat Commun* 8(1):1348. <https://doi.org/10.1038/s41467-017-01450-2>
7. R Core Team (2019) *R: a language and environment for statistical computing*, v. 3.5.3. R Foundation for Statistical Computing, Vienna. <https://www.R-project.org>
8. De Mendiburu F (2009) Una herramienta de análisis estadístico para la investigación agrícola. Master's thesis, Universidad Nacional de Ingeniería (UNI), Lima, Perú
9. Atwood C (1969) Optimal and efficient designs of experiments. *Ann Math Stat* 40: 1570–1602
10. He J, Li J, Huang Z, Zhao T et al (2015) Composite interval mapping based on lattice design for error control may increase power of quantitative trait locus detection. *PLoS One* 10(6):1–14. <https://doi.org/10.1371/journal.pone.0130125>
11. Tamhane A (1996) Design and analysis of experiments for statistical selection, screening and multiple comparisons. *Technometrics* 38(3):289–290. <https://doi.org/10.1080/00401706.1996.10484514>
12. Gadbury G, Garrett K, Allison D (2009) Challenges and approaches to statistical design and

- inference in high-dimensional investigations. *Methods Mol Biol* 553:181–206. https://doi.org/10.1007/978-1-60327-563-7_9
13. Liang Z, Qiu Y, Schnable J (2012) Distinct characteristics of genes associated with phenome-wide variation in maize (*Zea mays*). *bioRxiv* (2019). <https://doi.org/10.1101/534503>
14. Patterson H, Williams E (1976) A new class of resolvable incomplete block designs. *Biometrika* 63(1):83–92. <https://doi.org/10.1093/biomet/63.1.83>
15. Xu Y, Qiu Y, Schnable J (2018) Functional modeling of plant growth dynamics. *Plant Phenome J* 1:170007. <https://doi.org/10.2135/tppj2017.09.0007>
16. Wright K (2019) Desplot: plotting field plans for agricultural experiments, v. 1.5. Corteva Agriscience, Johnston. <https://CRAN.R-project.org/package=desplot>
17. Wright K (2018) Agridat: agricultural datasets, v. 1.16. Corteva Agriscience, Johnston. <https://CRAN.R-project.org/package=agridat>

Part II

Novel Algorithms for HTP



High-Throughput Extraction of Seed Traits Using Image Acquisition and Analysis

Chongyuan Zhang and Sindhuja Sankaran

Abstract

Seed traits can easily be assessed using image processing tools to evaluate differences in crop variety performances in response to environment and stress. In this chapter, we describe a protocol to measure seed traits that can be applied to crops with small grains, including legume grains with little modification. The imaging processing tool can be applied to process a batch of images without human intervention. The method allows evaluation of geometric and color features, and currently extracts 11 seed traits that include number of seeds, seed area, major axis, minor axis, eccentricity, and mean and standard deviation of reflectance in red, green, and blue channels from seed images. Protocols or methods, including the one described in this chapter, facilitate phenotyping seed traits in a high-throughput and automated manner, which can be applied in plant breeding programs and food processing industry to evaluate seed quality.

Key words Seed phenotyping, Feature extraction, Image processing, Grains

1 Introduction

Seed traits such as number of seeds and seed size are directly associated with crop productivity [1]. In addition, traits such as seed size and color are also associated with seedling vigor [2–5] and biochemical composition [6–9], and are important quality traits. The economics of produce (e.g. green pea, chickpea) depends on these factors [10]. Seed traits are often evaluated as a part of breeding trials and/or to assess stress response [1, 11, 12].

Conventionally, seed traits are evaluated using sieve analysis, estimating hundred or thousand seed weight, and color charts/scales. These methods can be subjective, labor-intensive, and very limited on the amount of data that can be acquired. Imaging with processing tools offer simple, reliable technique to extract number of features with high accuracy and throughput. Moreover, image features such as uniformity in seed size (that can also contribute to uniform maturity under field conditions) or seed size distribution,

which usually cannot be easily assessed using conventional methods. In this work, we describe a simple technique for extracting seed traits for grain crops.

2 Materials

2.1 Imaging System

1. Detach the cover of the office scanner (e.g., Epson Perfection V39, Epson America, Inc., Long Beach, CA, USA), and clean the glass top with 70% ethanol solution to remove any dirt and stain.
2. Use cardboard box covered with black paint inside to prevent interference from external light source, and that can serve as a background for the images (*see* **Notes 4.1., 1 and 2**).
3. Connect the computer to the scanner via software provided by the scanner manufacturer to acquire images (*see* **Note 4.1., 3**).

2.2 Auxiliary Items

1. Place the reference standard (Spectralon® Diffuse Reflectance Standards, SRS-99-020, Labsphere Inc., North Sutton, NH, USA) to correct for image reflectance for color measurements during image processing (*see* **Note 4.2., 1**).
2. Use size standard such as US Dime (17.91 mm in diameter) to convert the seed size from number of pixels to measurement units (mm, cm) (*see* **Note 4.2., 2**).
3. Place blue tape marker (as blue is not found in seeds) next to other auxiliary items to prevent the scanner from switching to back-and-white mode automatically.

2.3 Seeds

1. Clean the seeds as much as possible (especially when the seeds are small) to calculate number of seeds and hundred or thousand seed weight accurately during image processing.
2. Use the scanner system for small seeds such as wheat, camelina, and quinoa (for bigger seeds, *see* **Note 4.1., 3**).

2.4 Software

Use customized image-processing algorithm developed in MATLAB® (2018, MathWorks Inc., Natick, MA, USA) to extract seed features (*see* **Note 4.3., 1**).

3 Methods

3.1 Image Acquisition

1. Place seeds on the glass top of the scanner with some space between seeds and the auxiliary items mentioned above and cover the scanner with cardboard box described above (*see* **Notes 4.4., 1 and 2**).

2. Acquire images of seeds using the scanner system with 600 dpi (dots per inch).
3. Save the images as 8-bit images in JPG format.

3.2 Image Processing

Use the customized image-processing algorithm to automatically analyze acquired images. The image-processing steps are as described below:

1. Identify the reference panel automatically from original image (Fig. 1a using a mask shown in Fig. 1b), and correct the image based on incident light conditions. This step is important for color analysis and not for shape or size analysis. The digital number of each channel (red, green, and blue) is corrected by multiplying each pixel with a correction factor ($255/\text{mean pixel value of reference panel for the corresponding channel}$) to calculate corrected digital number that represents reflectance.
2. Create a mask image to separate the foreground of each image (seeds, blue tape marker, and size reference coin) using a set of thresholds (e.g., $R > 40$, $G > 30$, $B > 10$ for quinoa), as shown in Fig. 1c (see **Note 4.5., 1**).
3. Apply morphological operations to remove noise and increase the quality of the mask. For example, filling the gaps/holes within objects of foreground, removing small objects that are not seed, etc., as shown in Fig. 1d.
4. Apply watershed operation to separate seeds that are connected to each other, as shown in Fig. 1d.
5. Before extracting features from seeds, convert seed size from number of pixels to millimeter using a measurement coefficient calculated based on the major axis of the reference object (e.g., coin size).
6. Remove the coin, the blue tape marker, and objects that are too small or big to be seeds, as shown in Fig. 1e.
7. Extract features from each seed and calculate the (trimmed) mean values of seed features that represent the sample (variety, treatment, etc.). Before calculating the mean values of seed features, the biggest and smallest 15% of seeds (in terms of area) are removed to reduce the influence from connected and broken seeds and debris. Currently, the image-processing algorithm extracts 11 features from each image such as number (calculated before trimming), area, major axis, minor axis, eccentricity, and mean and standard deviation of reflectance in red, green, and blue channels of seed images (see **Notes 4.5., 2–5**).
8. Export the overlapping images of detected seeds and corresponding original images for quality inspection, shown in Fig. 1f, and seed features extracted from all the images as Excel file for further statistical analysis.

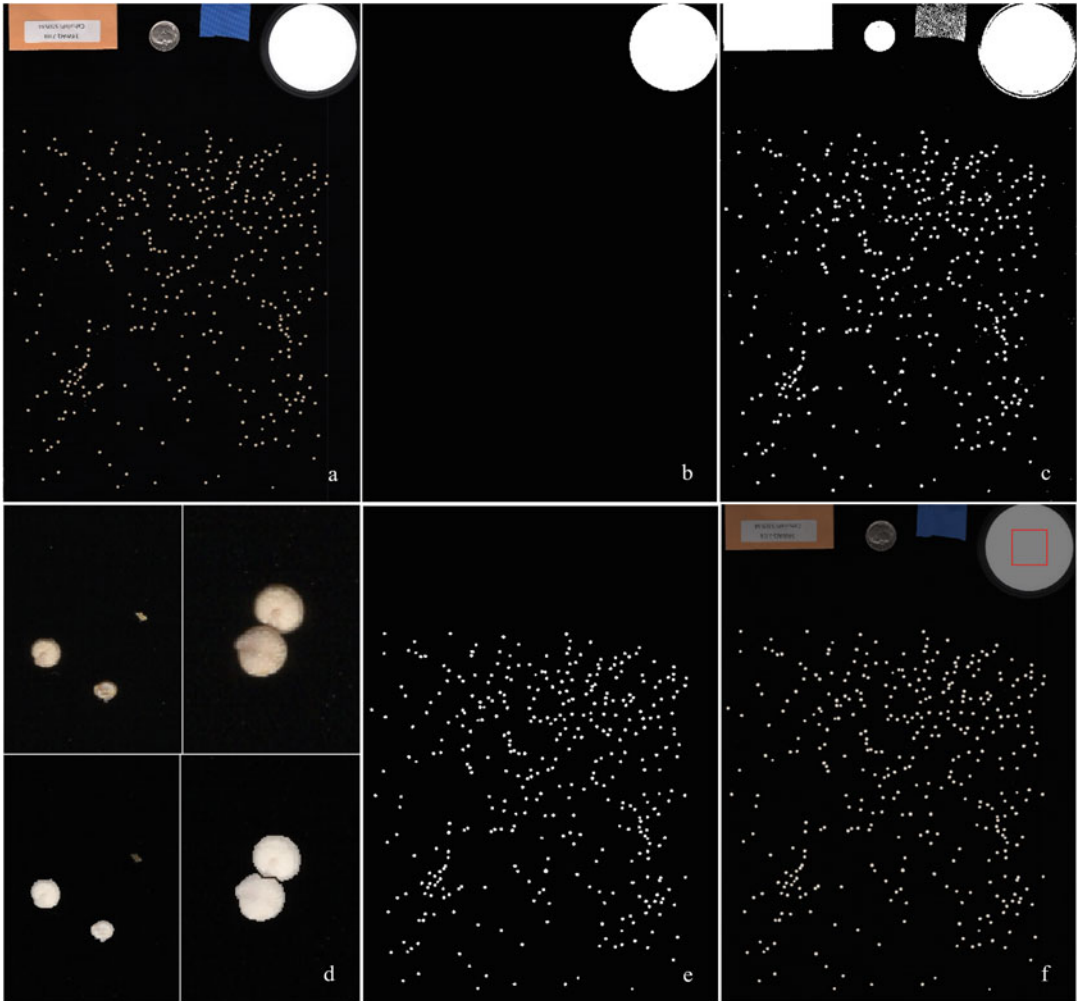


Fig. 1 Images showing processing procedure. Notes: (a) Original image of quinoa seeds and auxiliary items; (b) mask used to identify the reflectance panel; (c) mask used to separate foreground objects; (d) zoom-in images showing seeds/debris and connected seeds (the two on the top), and retained seeds (highlighted by white) with debris removed and seeds separated by watershed operation (the two at the bottom); (e) seed mask in which only seeds are retained; (f) overlapping of seed mask and original image, with seeds highlighted, while other objects are dimmed

4 Notes

4.1 Imaging System

1. Blocking external light source is important to create standardized condition for automated image processing and robust color feature extraction.
2. Select the right background to increase the efficiency of automated image processing. Background color will depend on the seed color. For dark colored seed, light backgrounds are recommended. In addition, coarse background can prevent specular reflection.

3. Phenotyping box similar as described in previous studies [13, 14] can also be used with customer-grade digital/RGB camera to phenotype bigger seeds, tubers, or other fruits/vegetables (cherry and pea). Similar applications can be found in literature [10, 15, 16].

4.2 Auxiliary Items

1. Recommended reference standard represents 99% reflectance from 250 to 2500 nm wavelength. However, other standards providing 99% reflectance or higher in visible range (380–740 nm) can be used.
2. Size standard can be of user choice, which should be relatively thin to avoid shadows and other image anomalies (2D object).

4.3 Software

1. The image-processing algorithm can be modified easily to adapt to different kinds of seeds and image resolution, to increase the features of interest to be extracted, etc. Image Processing Toolbox™ and Statistics and Machine Learning Toolbox™ in MATLAB are required to run this algorithm. Alternatively, ImageJ, PlantCV, GrainScan, SmartGrain, or others [10, 16–19] can be utilized for image processing and analysis.

4.4 Image Acquisition

1. Separating the seeds one by one during imaging is not necessary, as the image-processing algorithm can separate the seeds automatically. However, it is not recommended to place seeds as a clump, where dozens of seeds are touching or overlapped with each other.
2. It is suggested that the auxiliary items are places in a relatively fixed locations so that auxiliary items can be removed by masking these fixed locations. Detection and masking of auxiliary items automatically is possible, but it will slow down the processing speed.

4.5 Image Processing

1. The threshold may need to be adjusted based on seeds, background, and settings during imaging.
2. The standard deviation of reflectance in a channel is the mean of standard deviation of reflectance of the central 70% seeds after elimination of 15% on either directions as mentioned above (in terms of size/area) in a channel.
3. More color features such as mean and standard deviation of channels in HSV color space, Lab color space, and textural features can also be extracted by modifying the code [14].
4. Image-processing analysis can be adapted based on user needs.
5. MATLAB® code and sample images are available in Zenodo [20].

References

1. Komyshev E, Genaev M, Afonnikov D (2017) Evaluation of the SeedCounter, a mobile application for grain phenotyping. *Front Plant Sci* 7:1990
2. Ries SK, Everson EH (1973) Protein content and seed size relationships with seedling vigor of wheat cultivars. *Agron J* 65:884–886
3. Evans LE, Bhatt GM (1977) Influence of seed size, protein content and cultivar on early seedling vigor in wheat. *Can J Plant Sci* 57:929–935
4. Spilde LA (1989) Influence of seed size and test weight on several agronomic traits of barley and hard red spring wheat. *J Prod Agric* 2:169–172
5. Jahnke S, Roussel J, Hombach T et al (2016) phenoSeeder – a robot system for automated handling and phenotyping of individual seeds. *Plant Physiol* 172:1358–1370
6. Shirzadegan M, Röbbelen G (1985) Influence of seed color and hull proportion on quality properties of seeds in *Brassica napus* L. *Fette Seifen Anstrichm* 87:235–237
7. Kumar V, Rani A, Solanki S et al (2006) Influence of growing environment on the biochemical composition and physical characteristics of soybean seed. *J Food Compos Anal* 19:188–195
8. Li Y, Beisson F, Pollard M et al (2006) Oil content of *Arabidopsis* seeds: the influence of seed anatomy, light and plant-to-plant variation. *Phytochemistry* 67:904–915
9. Ayerza R (2010) Effects of seed color and growing locations on fatty acid content and composition of two chia (*Salvia hispanica* L.) genotypes. *J Am Oil Chem Soc* 87:1161–1165
10. Sankaran S, Wang M, Vandemark GJ (2016) Image-based rapid phenotyping of chickpeas seed size. *Eng Agric Environ Food* 9:50–55
11. Upadhyaya HD, Kashiwagi J, Varshney RK et al (2012) Phenotyping chickpeas and pigeonpeas for adaptation to drought. *Front Physiol* 3:179
12. Hinojosa L, Matanguihan JB, Murphy KM (2019) Effect of high temperature on pollen morphology, plant growth and seed yield in quinoa (*Chenopodium quinoa* Willd.). *J Agron Crop Sci* 205:33–45
13. Zhang C, Si Y, Lamkey J et al (2018) High-throughput phenotyping of seed/seedling evaluation using digital image analysis. *Agronomy* 8:63
14. Marzougui A, Ma Y, Zhang C et al (2019) Advanced imaging for quantitative evaluation of *Aphanomyces* root rot resistance in lentil. *Front Plant Sci* 10:383. <https://doi.org/10.3389/fpls.2019.00383>
15. Si Y, Sankaran S, Knowles NR et al (2017) Potato tuber length-width ratio assessment using image analysis. *Am J Potato Res* 94:88–93
16. Moore CR, Gronwall DS, Miller ND et al (2013) Mapping quantitative trait loci affecting *Arabidopsis thaliana* seed morphology features extracted computationally from images. *G3 (Bethesda)* 3:109–118
17. Whan AP, Smith AB, Cavanagh CR et al (2014) GrainScan: a low cost, fast method for grain size and colour measurements. *Plant Methods* 10:23
18. Tanabata T, Shibaya T, Hori K et al (2012) *SmartGrain*: high-throughput phenotyping software for measuring seed shape through image analysis. *Plant Physiol* 160:1871–1880
19. Gehan MA, Fahlgren N, Abbasi A et al (2017) PlantCV v2: image analysis software for high-throughput plant phenotyping. *PeerJ* 5:e4088. <https://doi.org/10.7717/peerj.4088>
20. Zhang C, Hinojosa L, Murphy K et al (2021) Seed color and size analysis using sample quinoa images. *Zenodo*. <https://doi.org/10.5281/zenodo.5752124>



ColourQuant: A High-Throughput Technique to Extract and Quantify Color Phenotypes from Plant Images

Mao Li, Margaret H. Frank, and Zoë Migicovsky

Abstract

Color patterning contributes to important plant traits that influence ecological interactions, horticultural breeding, and agricultural performance. High-throughput phenotyping of color is valuable for understanding plant biology and selecting for traits related to color during plant breeding. Here we present ColourQuant, an automated high-throughput pipeline that allows users to extract color phenotypes from images. This pipeline includes methods for color phenotyping using mean pixel values, a Gaussian density estimator of CIELAB color, and the analysis of shape-independent color patterning by circular deformation.

Key words Color phenotyping, High-throughput image acquisition, Color patterning, Continuous color distribution, Shape-independent color quantification

1 Introduction

Color patterning contributes to important ecological, horticultural, and agricultural traits. Developing high-throughput (HT) phenotyping methods for color analysis is essential for furthering our understanding of plant biology and providing accurate, quantitative information for plant breeding.

Morphological diversity in flower color plays a significant role in determining pollinator recruitment, and as a result, pollinator preference may result in selection for flower colors. For example, across 206 Australian angiosperm species, flowers pollinated by birds differed significantly in color from those visited by insects [1]. Pollinators may also exert selective pressure on color variation among close relatives, as evidenced by the hummingbird-pollinated clade *Iochrominae* (Solanaceae) [2].

In addition to driving plant-pollinator relationships, color is an essential component of the ornamental plant industry and has a direct influence on the commercial value of given cultivars [3, 4]. Indeed, the desire for new colors has been a major driver

behind the use of biotechnology in ornamental horticulture, especially for the cut flower industry [5].

Food color can also significantly influence flavor perception [6, 7], steering consumer choices. For fresh market produce, consumers prefer bright colors, which can indicate freshness and desirable nutrient content [8]. Novel fruit color may also add value to new fruit varieties. For example, in one study, 44% of consumers were willing to pay 50 cents more for a pear with red skin, despite its poor flavor [9]. However, traditional assessment of seedlings for color may be time-consuming for large numbers of plants and vary based on observer. Thus, efficient and accurate phenotyping of color and color patterning may serve as an important tool for plant breeding.

Quantitative color measurements may be used directly for culling plants without a desirable trait, or, in instances where the color is not apparent at the seedling stage (e.g., fruit on trees), the data can be used instead for genetic mapping. Techniques such as linkage mapping and genome-wide association studies (GWAS) connect phenotype data with genotype data to uncover genetic markers correlated with a trait of interest. Genetic mapping is improved by precise, quantitative data [10]. Early screening of plants using genetic markers allows the breeder to reduce the number of plants that is propagated without a trait of interest and therefore is especially cost-effective in perennial crops that have a lengthy juvenile phase, such as apples and grapes [11]. As a result, genomics-assisted breeding for color traits including peach blush [12] and sweet cherry fruit color [13] are already underway.

HT phenotyping can also be used to efficiently detect and diagnose pathogen spread in diseased plants [14], facilitating genetic mapping of disease resistance. Genomics-assisted breeding of disease resistance eliminates the need for the time-consuming and expensive task of inoculating plants. Digital imaging improves ease of scoring for infection and allows for a quantitative measurement of characteristics that would otherwise be missed. For example, a study of *Arabidopsis thaliana* infected with *Botrytis cinerea* reanalyzed images from a previous GWAS for visual traits, including color, finding that some resistance genes impacted color, but not the shape or size of lesions [15].

Among the benefits of HT color phenotyping is its potential to dramatically improve our characterization and understanding of plant diversity, such as plant-pollinator relationships, and have a direct impact on plant breeding for important traits including appearance and disease [4]. Here we present ColourQuant, methods for automated HT color phenotyping using mean pixel values, a Gaussian density estimator of CIELAB ($L^*a^*b^*$) color, and the analysis of shape-independent color patterning by circular deformation.

2 Materials

1. Flatbed scanner (e.g., Epson Perfection V550 Scanner) for flat images, camera, and light box for three-dimensional objects.
2. Color card (e.g., Kodak KOCSGS color separation guide).
3. MATLAB (<https://github.com/maoli0923/ColourQuant>).

3 Methods

3.1 Image Acquisition

1. For flat objects such as leaves, place samples on a flatbed scanner with a color card in the corner.
2. For three-dimensional objects, such as fruit, images may be acquired by placing the samples inside of a light box with a color card in the corner and photographing using a camera.
3. A few rounds of sample images should be collected in order to optimize lighting and resolution. It is useful to have enough light to capture details while reducing glare (*see Note 1*).
4. Color images can be saved in a variety of file formats. Lossless compression methods retain complete pixel information; these include TIF LZW and PNG file formats. JPG files are produced using Lossy compression, which reduces pixel information, making the files smaller, but less informative. Lossless compression is generally preferred for color image analysis; however, Lossy compression works in most cases and takes up less computer storage. The right file format will depend on the size of the experiment, hard drive space, and desired experimental output (*see Note 2*).
5. Color correct images. One method is to perform white balance for the image. In the example pictured, first extract and average R, G, and B values (denoting as avgR, avgG, and avgB) for the white swatch in the Kodak KOCSGS color separation guide. Then add $(255 - \text{avgR})$, $(255 - \text{avgG})$, and $(255 - \text{avgB})$ to R, G, and B of all the pixels in the image. As a result, the RGB value of the white swatch on the color guide is equal to 255 and the image color is white balanced (Fig. 1, *code lines 5–29*).

3.2 Object Segmentation

To separate the object (e.g., leaf, fruit) from the background, first convert the image into a binary image in which object and background are in white and black (or black and white), respectively. Many different segmentation methods exist, such as adaptive thresholding or learning algorithms [16]. In addition, parameters may need to be adapted based on factors such as brightness and contrast of the image with respect to the background, surrounding

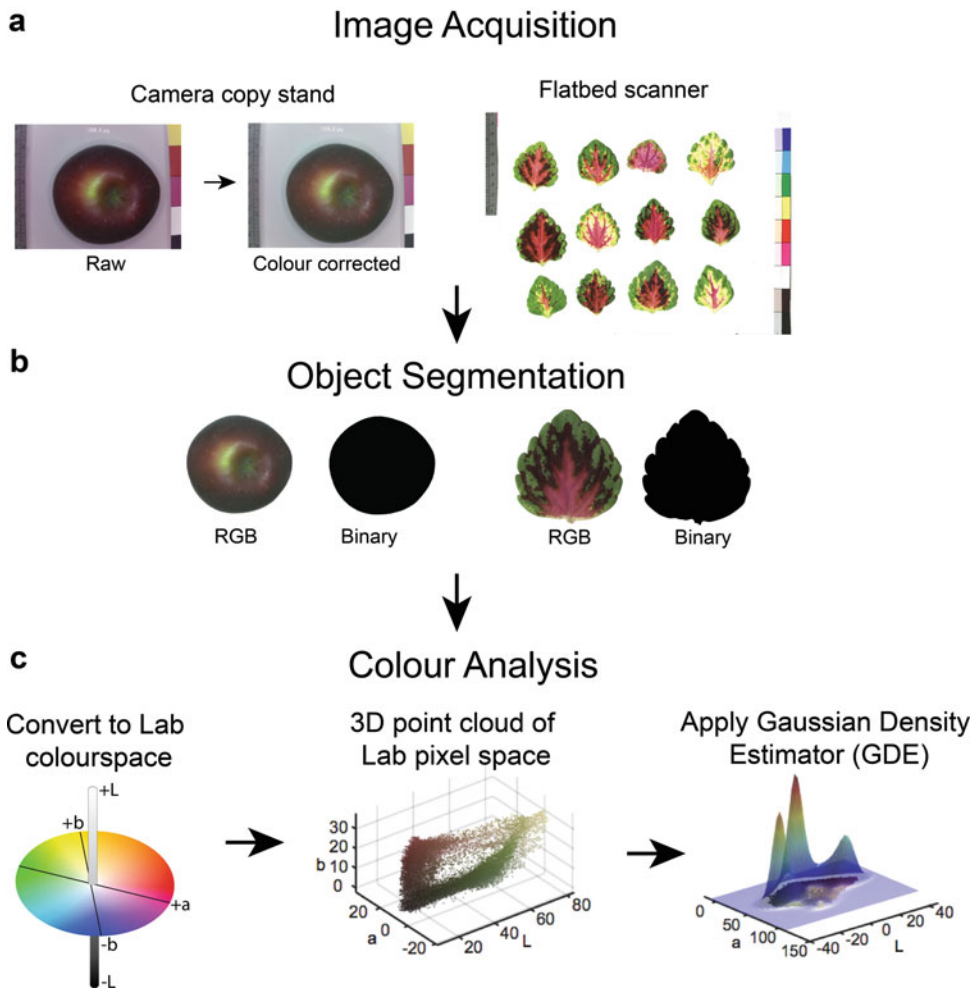


Fig. 1 Overview of ColourQuant pipeline going from image acquisition to global pixel quantification. Flat samples can be imaged using a scanner or digital camera, whereas three-dimensional samples are easiest to capture on a copy stand (a). For both approaches, the inclusion of a color card is essential for performing post-acquisition color balance. Color thresholding can be used to segment samples from the surrounding background (b), and then color values for isolated samples can be extracted and quantified using a three-step process. First, pixels are converted from RGB to $L^*a^*b^*$ continuous color space. This space is plotted in a three-dimensional point cloud and a Gaussian density estimator function is applied to the point cloud, in order to quantify the color composition of the sample (c)

shadows along the sample margins, and image quality. One object segmentation method that generally works well for scanned as well as photographed objects is outlined below (*code lines 31–69*) and displayed in Fig. 1.

1. Extract the RGB matrix from the image and convert it into hue-saturation-value (HSV) format.
2. In HSV, most background pixels become gray, and it is possible to set a threshold that separates gray values from true object

values. In our example, this value was Saturation > 0.15 , but it will need to be customized for each experiment (*code line 39*).

3. If there are multiple objects in an image (e.g., coleus image in Fig. 1), detect each connected component and segment out the object of interest (e.g., large enough objects, *code lines 44–51*) for color analysis.

3.3 Color Analysis

The first step in color analysis described here is to convert the color matrices from RGB to $L^*a^*b^*$ color (*code lines 78–85*). $L^*a^*b^*$ color is a continuous color space that consists of three descriptors: L^* = “lightness,” a^* = “green to magenta,” and b^* = “blue to yellow,” displayed in Fig. 1.

3.3.1 Mean and Variance

For objects with nearly solid colors or relatively simple patterns, calculating the mean and variance for L^* , a^* , b^* color values for each image is informative (*code lines 72–91*). These data can be summarized visually using a scatterplot with each object (e.g., apple fruit) displayed (Fig. 2).

3.3.2 Gaussian Density Estimator

For objects with complex color patterns, a robust and more comprehensive measurement tool, such as a Gaussian density estimator (GDE), needs to be applied.

1. Treat 3D $L^*a^*b^*$ color matrices as 3D point clouds with coordinates (L^* , a^* , b^*).
2. To reduce the amount of time needed for the computation, we find the extreme values for L^* , a^* , and b^* through 3D point clouds of the population (*code lines 88–90*). Then working 3D space can be bounded by a box with ranges based on the extreme values L^* , a^* , and b^* . In our example, L^* ranges between -10 and 110 , a^* ranges between -40 and 50 , and b^* ranges between -30 and 74 (*code lines 94–97*).
3. Extract color distribution and frequency for each image by applying a GDE to the $L^*a^*b^*$ point cloud (*code lines 98–112*). The GDE directly estimates density from the point cloud data; thus, it is a function defined on a 3D space (depicted in Fig. 1).
4. The GDE descriptor captures statistical color distributions; however, it does not provide information regarding spatial patterning. To capture spatial color information, the object can be segmented into distinct zones. One method is to define these zones based on normalized pixel distances (*code lines 114–136*), for example, the “border,” defined as the outer 15% of pixels from the leaf boundary to the centroid; the “center,” defined as the inner 75% of pixels from the centroid to the boundary; and “full,” defined as the entire color matrix. These zones should be customized for each study (*code lines*

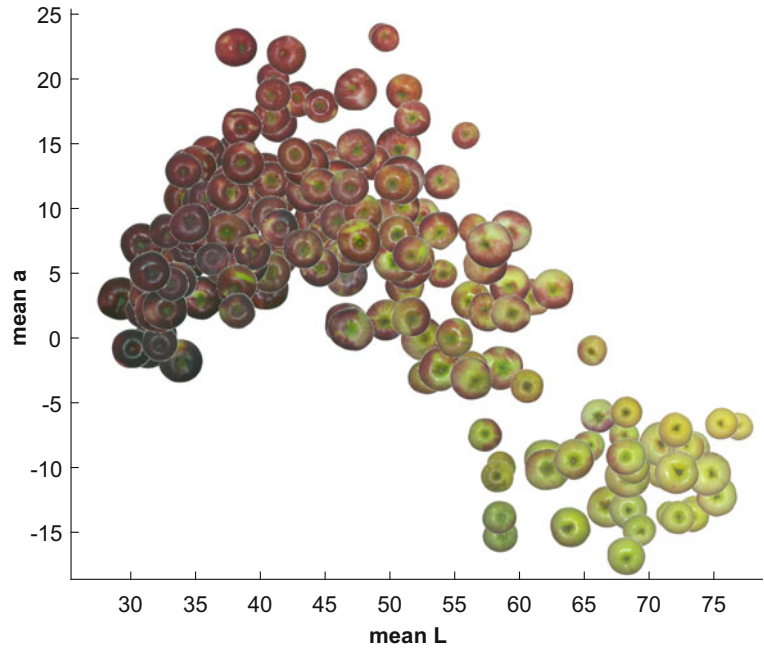


Fig. 2 Mean “L*” and “a*” pixel values displayed as a scatterplot of apple samples. In continuous color space, L* represents the dark-to-light spectrum of values and a* represents green-to-magenta space. In this example, apple samples spread from light green to deep red across the scatterplot

135–136). The distance between any two objects is calculated by

$$D = \sqrt{d_{\text{full}}^2 + d_{\text{border}}^2 + d_{\text{center}}^2}$$

where d represents the L_2 distance (the square root of the sum of the squared vector values) between GDE functions for each corresponding zone. This calculation determines the difference in color patterns between two objects based on their similarity across all zones (*code lines 152–156*).

3.3.3 Circular Deformation

To examine the impact of color while reducing the effect of shape, it is possible to deform each object (e.g., coleus leaf) into a disk using thin plate spline (TPS) interpolation [17]; algorithm is from [18].

1. Align all the objects to the same orientation (e.g., rotating the leaf so that the tip is on the top and base is on the bottom). This could be achieved by aligning a few manually or automatically labeled landmarks (e.g., leaf tip and base, *code lines 163–176*).
2. Normalize the object so that the square root of the average squared distances of all the points on the outline to the center is 1 (*code lines 177–178*).

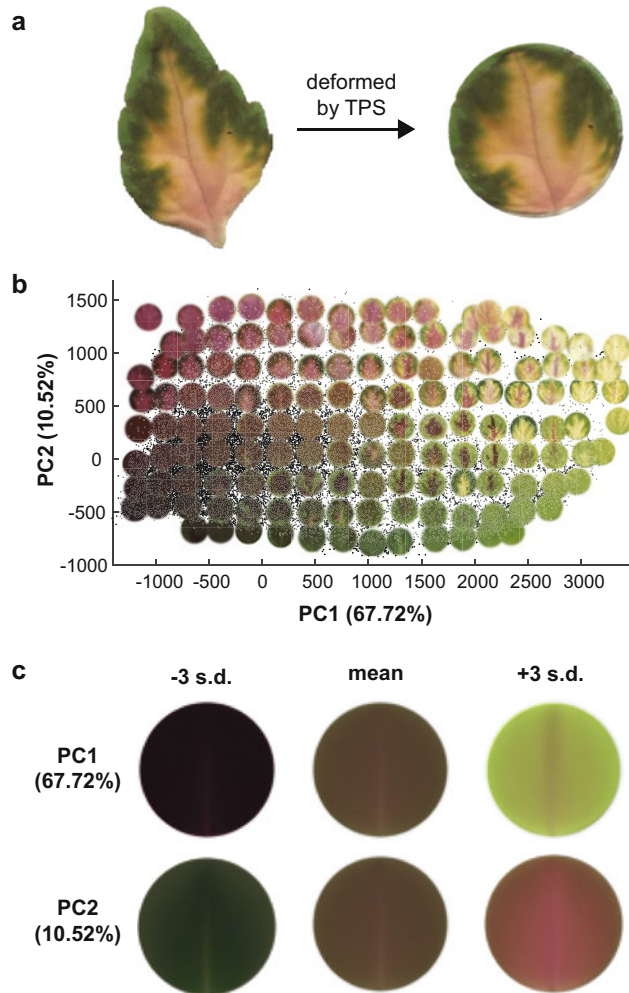


Fig. 3 Example of the thin plate spline method applied to coleus leaves to analyze color composition independent of sample shape. In this example, coleus leaves are deformed into circles using the thin plate spline method (a), and then plotted into PCA space based on their global pixel composition (b). Eigen leaves that explain the largest contribution to color variance across the sample population can be extracted from this analysis (c)

3. Set the points on the outline as control points.
4. Use TPS to deform the object to force the control points to be the points on the circle with radius 1 (Fig. 3a) and save the image with a transparent background (*code lines 182–199*).
5. Resize the circular image with a fixed dimension (e.g., 70×70). Extract L*a*b* colors from a 14,700-dimensional vector (4900 pixels, each has three values, *code lines 202–212*).

6. It is possible to perform principal component analysis on these vectors (Fig. 3b) and get eigen colors (Fig. 3c) that show color pattern variation (*code lines 214–232*).

4 Notes

1. Minimizing shadows around the edges of your samples will also help with streamlining the downstream image processing steps.
2. For large experiments, we recommend testing whether a Lossy file format (e.g., .jpg format) is appropriate for your study, or investing in additional data storage.

5 Conclusions

ColourQuant enables the efficient extraction of quantitative color distribution and patterning from a large set of samples. It can be applied to virtually any subject, and flexibly adapted to study different color patterns and investigate color patterning irrespective of sample shape. The output from this method includes a table of color values that correspond to each sample or subsample, for which there are numerous visualization and statistical packages built in R and MATLAB that can be used to analyze and plot the data.

Acknowledgments

Z.M. was supported by National Science Foundation (NSF) Plant Genome Research Program 1546869. M.H.F. is supported through startup funds from Cornell University's College of Agriculture and Life Science and by the NSF (CAREER IOS-1942437). The coleus samples used in Figs. 1 and 3 were imaged from Dr. David Clark's breeding program at the University of Florida in Gainesville, FL.

References

1. Shrestha M, Dyer AG, Boyd-Gerny S (2013) Shades of red: bird-pollinated flowers target the specific colour discrimination abilities of avian vision. *New Phytol* 198:301–310
2. Muchhala N, Johnsen S, Smith SD (2014) Competition for hummingbird pollination shapes flower color variation in Andean Solanaceae. *Evolution* 68:2275–2286
3. Zhao D, Tao J (2015) Recent advances on the development and regulation of flower color in ornamental plants. *Front Plant Sci* 6:261
4. Li M, Coneva V, Robbins KR et al (2021) Quantitative dissection of color patterning in the foliar ornamental coleus. *Plant Physiol* 187: 1310–1324
5. Chandler SF, Lu C-Y (2005) Biotechnology in ornamental horticulture. *In Vitro Cell Dev Biol Plant* 41:591

6. Spence C (2015) On the psychological impact of food colour. *Flavour* 4:21
7. Maga JA (1974) Influence of color on taste thresholds. *Chem Senses* 1:115–119
8. Barrett DM, Beaulieu JC, Shewfelt R (2010) Color, flavor, texture, and nutritional quality of fresh-cut fruits and vegetables: desirable levels, instrumental and sensory measurement, and the effects of processing. *Crit Rev Food Sci Nutr* 50:369–389
9. Gamble J, Jaeger SR, Harker FR (2006) Preferences in pear appearance and response to novelty among Australian and New Zealand consumers. *Postharvest Biol Technol* 41: 38–47
10. Burghardt LT, Young ND, Tiffin P (2017) A guide to genome-wide association mapping in plants. *Curr Protoc Plant Biol* 2:22–38
11. Luby JJ, Shaw DV (2001) Does marker-assisted selection make dollars and sense in a fruit breeding program? *HortScience* 36: 872–879
12. Sandefur P, Frett T, Clark J et al (2017) A DNA test for routine prediction in breeding of peach blush, Ppe-Rf-SSR. *Mol Breed* 37:11
13. Sandefur P, Oraguzie N, Peace C (2016) A DNA test for routine prediction in breeding of sweet cherry fruit color, Pav-Rf-SSR. *Mol Breed* 36:1–11
14. Mutka AM, Bart RS (2014) Image-based phenotyping of plant disease symptoms. *Front Plant Sci* 5:734
15. Fordyce RF, Soltis NE, Caseys C et al (2018) Digital imaging combined with genome-wide association mapping links loci to plant-pathogen interaction traits. *Plant Physiol* 178: 1406–1422
16. Yogamangalam R, Karthikeyan B et al (2013) Segmentation techniques comparison in image processing. *Int J Eng Technol* 5:307–313
17. Bookstein FL (1989) Principal warps: thin-plate splines and the decomposition of deformations. *IEEE Trans Pattern Anal Mach Intell* 11:567–585
18. Li M, Cole JB, Manyama M et al (2017) Rapid automated landmarking for morphometric analysis of three-dimensional facial scans. *J Anat* 230:607–618



Using Cameras for Precise Measurement of Two-Dimensional Plant Features: CASS

Amy Tabb, Germán A. Holguín, and Rachel Naegele

Abstract

Images are used frequently in plant phenotyping to capture measurements. This chapter offers a repeatable method for capturing two-dimensional measurements of plant parts in field or laboratory settings using a variety of camera styles (cellular phone, DSLR), with the addition of a printed calibration pattern. The method is based on calibrating the camera using information available from the EXIF tags from the image, as well as visual information from the pattern. Code is provided to implement the method, as well as a dataset for testing. We include steps to verify protocol correctness by imaging an artifact. The use of this protocol for two-dimensional plant phenotyping will allow data capture from different cameras and environments, with comparison on the same physical scale. We abbreviate this method as CASS, CAmera Scanner.

Key words Camera calibration, Image measurement, Plant phenotyping

1 Introduction

Images are used with increasing frequency in plant phenotyping for a variety of reasons. One reason is the ability to remotely capture data without disturbing the plant material, while another is the promise of high-throughput phenotyping via image processing pipelines such as those enabled by PlantCV [3]. However, to acquire precise data suitable for measurements of two-dimensional objects, the prevailing method in the community is to use a flatbed scanner. Shape analysis of leaves has used scanned images for apple, grapevine, *Claytonia* L., and a mixture of species [6–8, 11]. Scanners have also been used to analyze the shape of pansy petals [14] and *Vitis vinifera* L. seeds [9].

Cameras have been used to phenotype a range of structures and sizes, such as cranberry fruit shape and size [2] and root system architecture [1]. In both of these works, a disk of known diameter is added to the scene for scaling purposes.

1.1 Camera Calibration

The protocol in this paper transforms images acquired from a standard consumer camera such that measurements in pixels are representative of a planar scene. What this means in more detail is that we have emulated a flatbed scanner with a consumer camera; angles between lines are preserved, as are distance ratios. Physical measurements can be recovered from image measurements by dividing by the number of pixels per millimeter, similar to flatbed scanners.

This method is needed because measurements of two-dimensional objects, when done in image space of camera-acquired images, are subject to diminished accuracy from physical perturbations. A small movement of the camera up or down will give the erroneous impression that an object is larger or smaller in terms of pixels. Image pixels are also subject to radial distortion and projective geometry that allows three-dimensional objects to be viewed in a two-dimensional image. In other words, 100 pixels on one side of the image may not represent the same physical dimensions as 100 pixels in another portion of the image.

The method at the center of this protocol makes use of established camera calibration procedures to mitigate the problems of the preceding paragraph. Camera calibration is the estimation of parameters that relate three coordinate systems, image, camera, and world, to each other. Hartley and Zisserman [5] is a good text on camera calibration. When camera calibration is completed, the coordinate systems have been defined relative to a standard, and the relationships of one coordinate system to another are known.

Calibration patterns are used to define coordinate systems relative to a standard. These may take many forms; in this work we use aruco patterns [4]; laid out in a grid, patterns define the X–Y plane of the world coordinate system as in Fig. 1. The camera captures an image of the pattern to aid in defining the world coordinate system with respect to the image and camera coordinate systems.

Usually, many views of the pattern are captured to solve an optimization problem to fully calibrate the camera [15]. However, the structure from motion (SfM) community [10, 13] began exploiting EXIF data, or exchangeable image file format. EXIF data is a type of metadata that is common in today’s consumer cameras. Within SfM, the camera’s sensor size and some data from the EXIF file are used to generate an initial solution for some of the camera calibration parameters. We have borrowed this practice for calibrating in the phenotyping context.

1.2 Using a Camera as a Scanner

The original intent of this method was to develop a high-throughput substitute for slow flatbed scanners. The steps in Sub-heading 3 will give details for the user. A brief overview of the code is provided with this chapter: (1) calibrates the camera, per image, (2) computes the homography to transform the current image to the X–Y grid of the world coordinate system, and (3) warps the current image to match the world coordinate system’s X–Y grid.

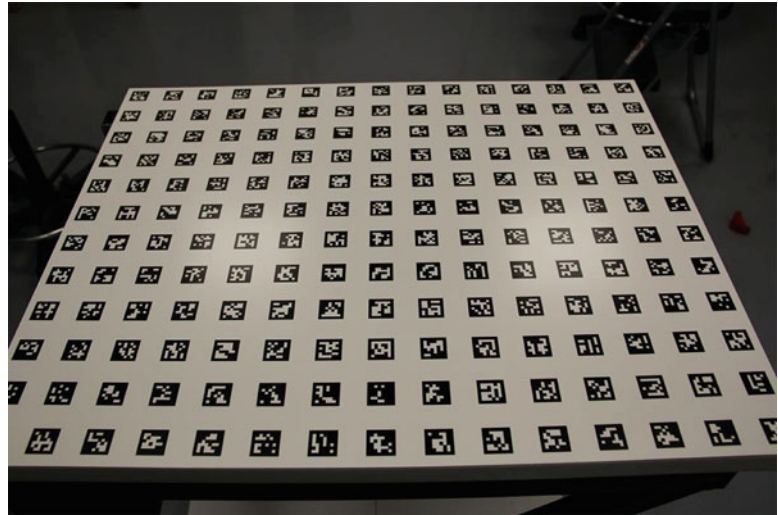


Fig. 1 Example of a grape cluster. This is a three-dimensional object, but we are interested in measuring aspects of the object where it meets the calibration pattern. Top row: input images of the same grape cluster, left two images are from an Apple iPhone 6 (cellular phone camera), right two images are from a Canon EOS 60D DSLR camera. Bottom row: results of applying the method for the image above, where every 10 pixels equal 1 mm. Full images are available in camera-as-scanner data

Figure 2 shows the input images and the output of the method. From the output images, users can apply their own computer vision techniques to identify the objects of interest. Measurements in pixels can be transformed to physical units by dividing by the user-selected scaling factor.

It is important to note a strong assumption when using this method, which is that the object is planar. In practical terms, the user should either use objects that are roughly planar or consider the footprint of the object on the calibration pattern plane. This method is not suitable for measuring objects that are nonplanar, such as freestanding branches with the calibration pattern behind.

To verify that the protocol has been performed correctly, we also include instructions for verifying that the measurements are correct by way of an artifact.

2 Materials

The materials needed are:

1. Calibration pattern.
2. Camera.
3. Artifact.
4. Code.

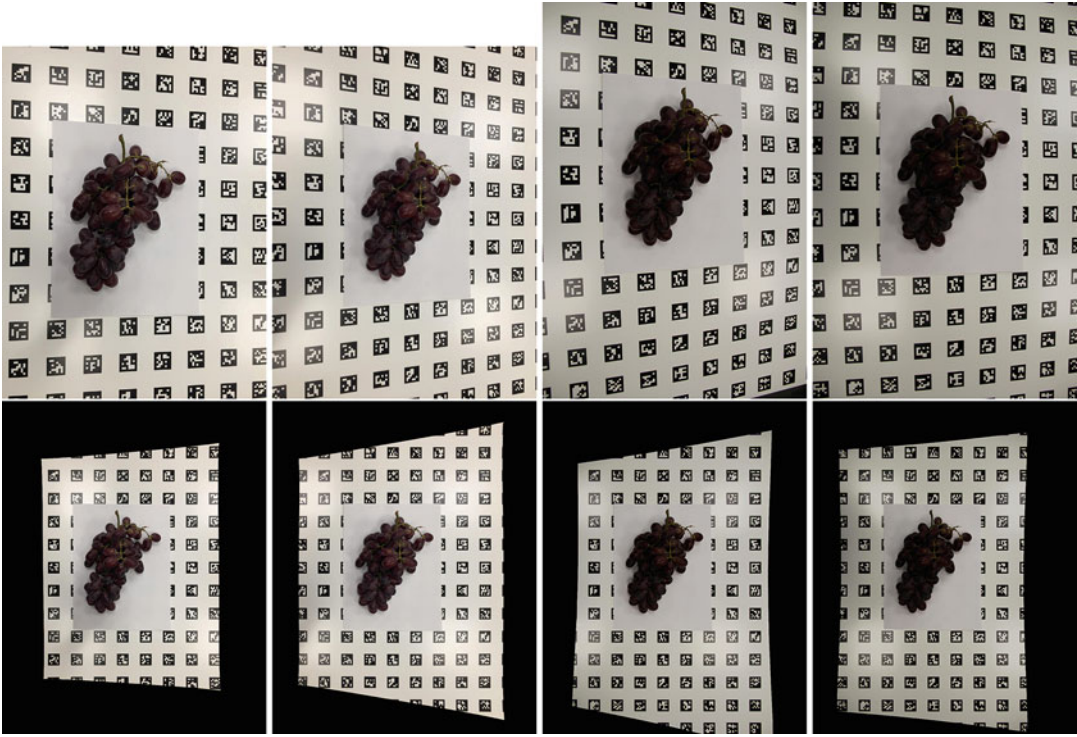


Fig. 2 An aruco calibration pattern. This particular example has been printed on aluminum, so it can be cleaned during experiments, which is convenient in plant research

The preparation of the calibration pattern is documented in **step 1**, Subheading **3**. The style of the camera is not specific to this method, and should be chosen for the user's convenience. This method relies on the extraction of EXIF tags, so the camera should write EXIF data. At the time of this writing, this feature is common in consumer and cellular phone cameras. An artifact of a known size is needed to check that the protocol has been implemented correctly. In our example, we chose a playing card, as shown in Fig. 3. A natural choice for an artifact may be a ruler.

The code and test datasets are provided in [12]. Within [12] are two programs and the data source: `aruco-pattern-write`, `camera-as-scanner`, and `data camera-as-scanner` data. To prepare for the experiments, install the code and run the examples.

3 Methods

1. Prepare the aruco calibration pattern. The pattern should be printed such that x and y axes are equally scaled, and attached to a flat surface. A pattern is provided in the [12] resource, as well as code for generating a new pattern via `aruco-pattern-write` and instructions in its README. Considerations when

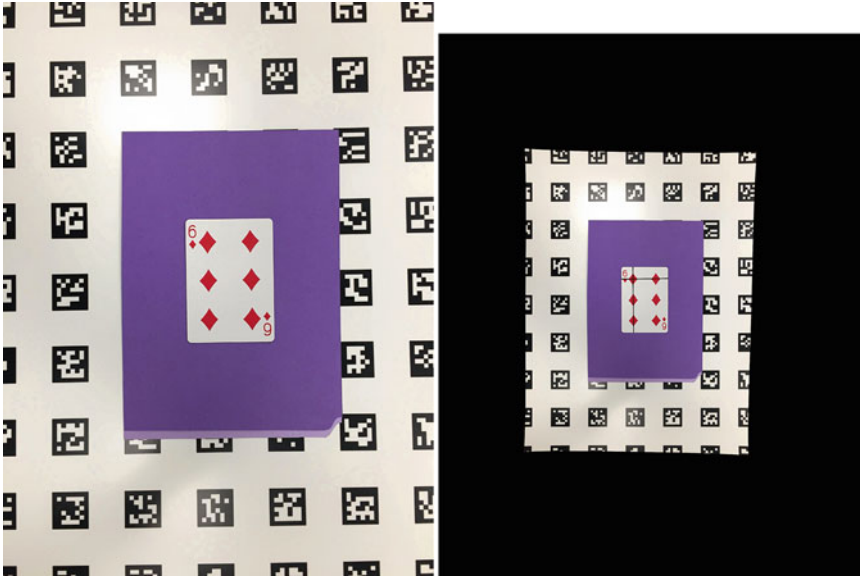


Fig. 3 Left: Apple iPhone6 camera images of a 2.5×3.5 inch ($63.5 \text{ mm} \times 88.9 \text{ mm}$) playing card. Right: results of applying the method for the image above, where every 10 pixels equal 1 mm. Black lines indicate measurements of the card in pixels. The horizontal line was 635.02 pixels, so is equivalent to 63.502 mm as measured by this system. The vertical line was 888.07 pixels, which is equivalent to 88.807 mm

generating a new pattern are in **Note 1**. The option of printing patterns on metal is discussed in **Note 2**.

2. Arrange the object to be measured on top of the aruco pattern printout. If segmentation of the object from the scene is desired using an image processing technique, we suggest placing a solid-colored paper or fabric in between the object and the pattern. *See Note 3* for more details.
3. Acquire images of the object, including at minimum a one-layer border of aruco tags on all four sides of the image. The image should generally be in focus, and acquired such that the camera body is parallel to the aruco pattern plane. However, the alignment does not have to be exact. *See Figs. 2 and 3* for examples. If using a cell phone camera, do not zoom. Standard image formats are all acceptable, as long as EXIF tags are generated.
4. Acquire an image of an artifact (such as a ruler) of known size with the same protocol as in **step 3**. We suggest that the artifact be rectangular in shape to allow for ease of measurement.
5. Prepare the image and format information to run camera-as-scanner. This step assumes that the code has been installed according to its instructions, mentioned in Subheading 2.

- 5.1 The preparation instructions for running the method for a group of images are given with the README of repository camera-as-scanner. Create a test directory.
- 5.2 Look up the camera's sensor size and convert to millimeters. This information may be found in the manufacturer's provided information that came with the camera, or can be found online. Fill in the sensor size parameters in the appropriate file as indicated in **step 5.1**.
- 5.3 Measure one of the squares of the printed aruco calibration pattern, in millimeters. Fill in the square length parameter of the appropriate file as indicated in **step 5.1**.
- 5.4 Move the images of the objects and image of the artifact to a directory with the name *images* within the test directory.
- 5.5 Determine the number of pixels per millimeter $np \in (0, \infty)$ for the transformed images, which will be an argument for running the code. The choice for np depends on the size of the object, size of the calibration pattern, and how large one can tolerate the result image size. Suppose the aruco calibration pattern print is x mm $\times$ y mm. The result images will be $x * np$ pixels $\times$ $y * np$ pixels. See **Note 4** for suggestions. In Figs. 2 and 3, 10 was chosen.
6. Run the code camera-as-scanner with three, and optionally four, arguments: the directory and the specified files and directory from **step 5**, an empty output directory, and the number of pixels per millimeter np . The optional fourth argument is a Boolean variable, 0 or 1, indicating whether intermediate results are written. If the variable is 1, the intermediate results are written; if 0, they are not.
7. Verify that the output is as expected, by inspecting the warped image corresponding to the artifact. Measure the width of the artifact in an image manipulation program such as ImageJ, KolourPaint, the GIMP, Adobe Photoshop, etc.; its units will be pixels w_p . Measure the width of the physical artifact in millimeters: w_{mm} . The following should be true: $w_{mm} = \frac{w_p}{np}$. If not, then recheck the steps. The verification process was demonstrated with the playing card artifact in Fig. 3.

4 Notes

1. Note that the pattern can be scaled up or down to be suitable for the data acquisition context, such as the image provided in aruco-pattern-write as an example. It is not necessary for the camera to view the whole pattern. The patterns are black and white, so do not need to be printed in color.

2. In our experiments, we have ordered prints of the patterns on aluminum. These have been convenient when working with fruit and plant material, because aluminum prints can be washed and cleaned. It is important that the aruco patterns not become occluded with dirt or stains.
3. Concerning segmentation of the object from the scene of aruco pattern and solid-colored fabric or paper, we suggest that the solid-colored fabric or paper be chosen such that it is a contrasting color compared to the target object. The fabric or paper should be cleaned or replaced if there are dirt or stains. The color of the fabric or paper, whatever color is chosen, will not interfere with the detection of the aruco tags.
4. As np increases, so will the image size. We suggest trying a range of sizes with a small number of images, such as $np = 5, 10, 20$, to get a sense of the resulting file size and resolution of features of interest.

Acknowledgments

We acknowledge the support of USDA-NIFA Specialty Crops Research Initiative, VitisGen2 Project (award number 2017-51181-26829). A. Tabb is supported by USDA-ARS Project 8080-21000-032-00-D.

References

1. Das A, Schneider H, Burrige J et al (2015) Digital imaging of root traits (DIRT): a high-throughput computing and collaboration platform for field-based root phenomics. *Plant Methods* 11(1):51. <https://doi.org/10.1186/s13007-015-0093-3>
2. Diaz-Garcia L, Covarrubias-Pazaran G, Schlautman B et al (2018) Image-based phenotyping for identification of QTL determining fruit shape and size in American cranberry (*Vaccinium macrocarpon* L.). *PeerJ* 6:e5461. <https://doi.org/10.7717/peerj.5461>
3. Fahlgren N, Feldman M, Gehan M et al (2015) A versatile phenotyping system and analytics platform reveals diverse temporal responses to water availability in *Setaria*. *Mol Plant* 8(10): 1520–1535. <https://doi.org/10.1016/j.molp.2015.06.005>
4. Garrido-Jurado S, Munoz-Salinas R, Madrid-Cuevas F et al (2014) Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recogn* 47(6): 2280–2292. <https://doi.org/10.1016/j.patcog.2014.01.005>
5. Hartley R, Zisserman A (2003) Multiple view geometry in computer vision, 2nd edn. Cambridge University Press, New York
6. Klein LL, Caito M, Chapnick C et al (2017) Digital morphometrics of two North American grapevines (*Vitis: Vitaceae*) quantifies leaf variation between species, within species, and among individuals. *Front Plant Sci* 8:373. <https://doi.org/10.3389/fpls.2017.00373>
7. Li M, An H, Angelovici R et al (2018) Topological data analysis as a morphometric method: using persistent homology to demarcate a leaf morphospace. *Front Plant Sci* 9:553. <https://doi.org/10.3389/fpls.2018.00553>
8. Migicovsky Z, Li M, Chitwood DH et al (2018) Morphometrics reveals complex and heritable apple leaf shapes. *Front Plant Sci* 8: 2185. <https://doi.org/10.3389/fpls.2017.02185>
9. Orr M, Grillo O, Lovicu G et al (2013) Morphological characterization of *Vitis vinifera*

- L. seeds by image analysis and comparison with archaeological remains. *Veg Hist Archaeobot* 22(3):231–242. <https://doi.org/10.1007/s00334-012-0362-2>
10. Snavely N, Seitz SM, Szeliski R (2006) Photo tourism: exploring photo collections in 3D. In: ACM SIGGRAPH 2006 papers, SIGGRAPH '06. ACM, New York, pp 835–846. <https://doi.org/10.1145/1179352.1141964>
 11. Stoughton TR, Kriebel R, Jolles DD et al (2018) Next-generation lineage discovery: a case study of tuberous *Claytonia* L. *Am J Bot* 105(3):536–548. <https://doi.org/10.1002/ajb2.1061>
 12. Tabb A (2020) Data and code from: using cameras for precise measurement of two-dimensional plant features: CASS (v 1.0). Zenodo. <https://doi.org/10.5281/zenodo.3677473>
 13. Wu C (2013) Towards linear-time incremental structure from motion. In: 2013 international conference on 3D vision. IEEE, Seattle, pp 127–134. <https://doi.org/10.1109/3DV.2013.25>
 14. Yoshioka Y, Iwata H, Hase N et al (2006) Genetic combining ability of petal shape in garden pansy (*Viola wittrockiana* Gams) based on image analysis. *Euphytica* 151(3):311–319. <https://doi.org/10.1007/s10681-006-9151-2>
 15. Zhang Z (2000) A flexible new technique for camera calibration. *IEEE Trans Pattern Anal Mach Intell* 22(11):1330–1334. <https://doi.org/10.1109/34.888718>

Part III

Molecular Plant Imaging



Chapter 11

Positron Emission Tomography (PET) for Molecular Plant Imaging

Sergey Komarov and Yuan-Chuan Tai

Abstract

Positron emission tomography (PET) is an imaging technology that measures 3D spatial distribution and kinetics of radio-tagged biomolecules in a living subject quantitatively and nondestructively. Commonly used positron-emitting radionuclides include ^{11}C , ^{13}N , and ^{15}O , which are essential elements for plant growth. Combining radiotracer techniques with PET, this in vivo molecular imaging capability offers plant biologists a powerful tool for molecular phenotyping research. While PET is widely used clinically for cancer diagnosis and pre-clinically for drug development, it is an unfamiliar imaging tool for plant biologists. This chapter introduces the basic principles of PET, factors that affect the quantitative accuracy of PET when imaging plants, and techniques for administering radiotracers to plants for a variety of molecular plant imaging applications.

Key words Positron emission tomography, PET, Molecular imaging, Plant phenotyping

1 Introduction

A wide variety of imaging technologies have been adapted by plant scientists to provide structural and functional information of plants nondestructively [1, 2]. Depending on the technologies employed, some are suitable for high-throughput and detailed imaging analysis of a large number of samples in a controlled laboratory settings [3], while others can operate in the field to collect data from a natural growing environment [4, 5]. The longitudinal measurements of plants' morphology and physiology over their life cycles allow us to establish more accurate biological models to predict the growth and yield of plants, as well as how plants respond to their environments [6].

Among all the imaging technologies that measure the phenotypic characteristics of plants, positron emission tomography (PET) technique measures the spatial and temporal distribution of radio-tagged biomolecules in a whole plant quantitatively. The “tag”

employed here is a radionuclide that decays through the emission of a positron—the anti-matter of electron that carries the same mass but opposite charge of an electron. The emitted positron carries a significant amount of energy (typically in the range of hundreds of keV or more). It interacts with the surrounding medium to continuously slow down and eventually annihilate with an electron. The result of its annihilation is the instantaneous emission of two back-to-back gamma rays of 511 keV each (the rest mass of an electron or a positron). The detection of two 511 keV gamma rays simultaneously is the underlying principle of a PET scanner. One can use two planar detectors to sandwich an object to obtain a “projection image” of radioactivity distribution in a three-dimensional (3D) object. Most PET scanners employ one or more rings of detectors to surround an object and to collect coincidence events from multiple angles simultaneously. Through mathematical reconstruction algorithms of the coincidence events measured by multiple pairs of detectors, tomographic images can be derived to show the 3D distribution of the radioactivity within an object. Measurement of such information over time can reveal the kinetics of radiolabeled biomolecules in a subject, which is often used to establish biological models of interest (such as a disease model or the pharmacokinetics of a new drug).

The image resolution of PET is known to be limited by three fundamental factors: positron range, photon acolinearity, and the intrinsic resolution of detectors. Positron range refers to the distance between origin of the positron (where the radioactive decay of the mother nuclide takes place) and the origin of the detected signals (two annihilation gamma rays). This uncertainty depends on the kinetic energy of a positron when it is emitted from a nucleus. Depending on the type of radionuclide used to tag biomolecules, the level of blurring to image resolution can be as small as 200–300 μm or as large as several mm. The photon acolinearity refers to a small angular uncertainty between the two back-to-back annihilation gamma rays because the positron and electron are not completely at rest when they annihilate. To preserve the momentum, the two gamma rays cannot be exactly 180-degree apart. The level of uncertainty in localizing the origin of the two gamma rays depends on the distance between the pair of detectors. Therefore, the bigger the ring diameter, the more blurring there is to the image resolution. For a small-ring scanner that is ~ 10 cm in diameter, the blurring may be only 200 μm . For a clinical PET scanner that has 90 cm diameter detector rings, the blurring may be as high as 2 mm. In terms of detector intrinsic spatial resolution, it is often a balance between the cost-effectiveness and performance. The highly penetrating 511 keV gamma rays require substantial detector mass in order to interact with and stop the photons effectively. For this reason, detectors of larger dimension should be used to increase the probability of detection, but often at the expense of

reduced intrinsic spatial resolution. There are sophisticated radiation detectors that employ complex readout schemes to achieve submillimeter intrinsic spatial resolution in 3D. These may be useful when imaging small animals for preclinical or pharmaceutical research, but could be a waste of resource if used for a clinical PET scanner whose image resolution is fundamentally limited by the photon acolinearity to be no better than 2 mm regardless of the choice of detector technologies. As a result, the image resolution of a clinical PET scanner remains in the 3–4 mm range, while the image resolution of a small animal PET scanner for studying rodents is approximately 1 mm. Details on the basic physics of PET and its applications can be found in an excellent reference book here [7].

In this chapter, we will explain the basic components needed for PET imaging. We will point out unique challenges (relative to human and animal imaging), limitations, and solutions when applying PET imaging to study plants. We include several examples to illustrate the molecular imaging capability of PET and its potential for plant phenotyping.

2 Methods

2.1 Radiotracer and PET Imaging for Plants

PET has been widely used clinically for cancer staging, restaging, and evaluating the efficacy of cancer therapies (such as chemotherapy). The radiolabeled biomolecule commonly used to detect cancer is ^{18}F -FDG—a glucose analog tagged by ^{18}F to measure glucose metabolism in a patient's body. The choice of ^{18}F as a radioactive tag for clinical PET imaging is due to its half-life ($T_{1/2}$) of 109.8 min, which permits a radiopharmaceutical to be manufactured in a centralized production facility and distributed to hospitals within a radius of a couple hundred miles. Several other commonly used positron-emitting radionuclides such as ^{11}C , ^{13}N , and ^{15}O are the essential elements for plant growth. Therefore, it is unsurprising that ^{11}C was used to study plants well before PET was widely accepted and used clinically [8–10]. However, with the short half-lives of 20.33 min, 9.97 min, and 2.04 min for ^{11}C , ^{13}N , and ^{15}O , respectively, one will need to have access to a cyclotron or a linear accelerator nearby in order to produce these radioisotopes for PET imaging. The availability of radionuclides and the cost associated with a cyclotron operation are two of the major limiting factors for PET to be widely adapted as a phenotypic imaging tool to study plants. Fortunately, with PET now widely adapted for clinical imaging and preclinical research, many medical centers now have on-site cyclotron facility that can be leveraged to supply ^{11}C and/or ^{13}N to support plant imaging research. As a result, we have seen increased interest in the use of PET for in vivo mapping of molecular events in plants [11–14]. Throughout this

chapter, it is assumed that one would have access to the short-lived radionuclides such as ^{11}C or ^{13}N from a cyclotron facility nearby. Otherwise, specialized radiotracers labeled with longer-lived isotopes such as ^{18}F ($T_{1/2} = 109.8$ m) or ^{64}Cu ($T_{1/2} = 12.7$ h) may need to be used in order to overcome this limitation.

The relatively short half-life of ^{11}C and ^{13}N also limits the biological processes that can be observed using PET imaging. In general, the “window of opportunity” for observing a biological process using radiotracer imaging techniques is less than ten times of the radioactive decay half-life of the isotope. For ^{11}C , this is approximately 3 h. For ^{13}N , this is less than 100 min. While this may be a major limitation for observing slow biological process in plants, it is also an advantage for measuring dynamic processes and/or transient responses of plants to external perturbations. For example, a plant can be studied repetitively every 3 h using $^{11}\text{CO}_2$ as a label to probe its photosynthesis efficiency and carbon allocation in responses to environmental stimuli. This nondestructive measurement of transient molecular processes in plants is a unique strength of PET that is otherwise difficult to gain using stable isotopes or other imaging techniques. That said, the short half-lives of ^{13}N and (in particular) ^{15}O require careful planning and execution of the imaging study because a large amount of radioactivity needs to be administered in order to have a measurable signal at the end of the imaging experiments. It is advised to carefully plan and execute a PET imaging experiment with a “cold run.” That is, one should run through the entire experiment without using radioactive tracer to familiar oneself with all the essential steps and to identify potential errors in experimental design. This will help to minimize radiation exposure to personnel and to ensure the success of the experiment.

Most PET scanners have horizontal bore to accommodate human or animal imaging applications. Several groups developed application-specific PET scanners dedicated to plant imaging research, including the PETTIS [15], PlanTIS [16], PlantPET [13, 17], etc. The PlantPET scanner is mounted inside a plant growth chamber to enable imaging experiments under a controlled environment. Additionally, respired radioactive gases from plants will be exhausted to provide additional protection to workers during an experiment. Since a PET study can take up to a few hours, appropriate protection against radiation exposure should be factored in when designing a PET imaging lab for plant imaging studies. The control room of the scanner and/or other associated equipment may need to be lead-lined for door, walls, and glass windows to provide adequate shielding for 511 keV gamma rays.

2.2 Challenge in Quantitative Accuracy of PET Images When Imaging Plants

A unique challenge when applying PET to plant imaging applications is the escape of positrons from the object when the radiolabeled biomolecules are in small or thin structures of a plant surrounded by air. For example, $^{11}\text{CO}_2$ can be assimilated into a leaf through photosynthesis. A positron emitted from ^{11}C -labeled sucrose within a thin leaf or shoot may quickly escape the plant and travel a long range before it annihilates with the surrounding structures [18]. The escaped positrons induce false signals (annihilation gamma rays) in medium surrounding the plant. In contrast, the escaped positrons are typically less of an issue in human or animal PET imaging studies unless the focus of the study is the lungs where tissues are partially surrounded by air. When a significant fraction of positrons can escape and annihilate at locations far away from their origins, not only is the image resolution degraded but also more importantly the quantitative accuracy of PET images is compromised. This may not be a problem for root imaging where the escaped positrons can still annihilate with the surrounding soil. However, it could lead to less quantitative results when imaging leaves or small plant structures in air.

To address this challenge, one may wrap the tissue-of-interest (e.g., fine shoot structures or leaves) with a thin layer of medium to force the escaped positrons to annihilate near their origins. Figure 1a shows a leaf containing ^{11}C -labeled photosynthates sandwiched by two plastic plates to force the escaped positron to annihilate in the plastic plates. Figure 1b shows small stems surrounded by short and thin plastic tubes when one measures the ^{11}C -labeled photosynthates flowing from the (upstream) leaves to roots through phloem in these stems. Notice that the plastic plates surrounding the leaf in Fig. 1a are transparent so that they do not interfere with photosynthesis. Air gaps are also allowed between the medium applied (plastic plates or tubes) and the plant tissues to avoid the blocking of gas exchange between the leaves and air that might alter a plant's physiology and functions. Figure 2 shows a common bean plant with its roots in a hydroponic solution that contains ^{13}N -ammonia. Uptake of ^{13}N -ammonia was transferred upward to leaves. A small plastic tube around the stem forces all escaped positrons to annihilate, which significantly enhanced the intensity of PET images on the right.

Despite the challenge of “escaped positrons,” PET images of a plant containing radiotracers are at least semiquantitative because a significant fraction of emitted positrons has sufficiently low energies that will still annihilate in the plants to form PET images. The pixel value in these images is still proportional to the activity concentration except that annihilation efficiency may be dependent on the structural thickness of the plant. Before the size of a plant changes significantly, PET images of the same plant following multiple rounds of radiotracer labeling (such as $^{11}\text{CO}_2$) can be compared semiquantitatively to measure the relative changes of radiotracer uptake caused by environmental changes or biological effects.

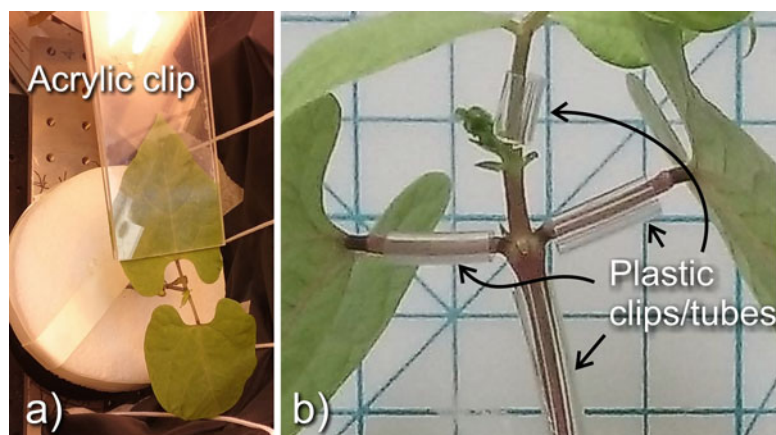


Fig. 1 To address the issue of escaped positrons from thin leaves and stems, plastic clips and tubes can be used to force annihilation: (a) Acrylic clip applied around a leaf after $^{11}\text{CO}_2$ was administered; (b) Plastic tubes applied around stems and petioles of a young common bean plant

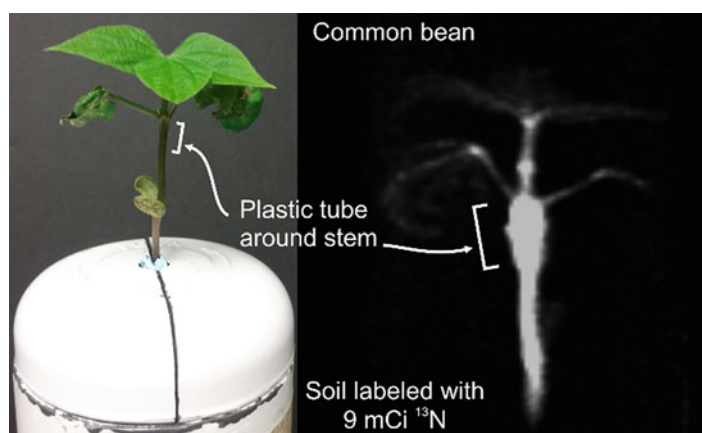


Fig. 2 A plant of common bean with the soil (enclosed in a lead container) labeled with ^{13}N -ammonia (333 MBq or 9 mCi) water solution. Plastic tube (not shown on the left photo) was applied around the stem to force escaped positrons to annihilate. Forced annihilation of escaped positrons increased PET image intensity at the location of the plastic tube

2.3 Administer Radiotracers to Plants

To perform PET imaging study of human or animals, radiotracers are commonly administered through intravenous injection, taking advantage of the blood circulation system to rapidly deliver radiotracer to the entire body. Other routes of administration include inhalation, ingestion, intraperitoneal injection, etc., depending on the type of biomolecules used and the physiological functions to be measured. Plants, on the other hand, have relatively slow circulation systems such as phloem and xylem. Direct injection of

radiotracer into either system is technically challenging. Excessive handling of a plant is also considered a significant stress to the plant itself and may alter the plant physiology and functions of interest. Tran et al. [19] have demonstrated ^{18}F -labeled sucrose can be introduced to a plant's phloem system through a cut at the tip of a leaf. Alternatively, $^{11}\text{CO}_2$ can be rapidly assimilated by a plant through photosynthesis to generate ^{11}C -labeled sucrose whose translocation in a plant through the phloem system can be tracked over time by PET imaging. Similarly, ^{13}N -labeled ammonia or nitrate [20] can be administered to a plant through its roots to follow their translocation through the xylem system.

Depending on the type of radiolabeled molecules used, the uptake of the radiotracer may be very efficient in some cases (e.g., plants are extremely efficient in assimilating $^{11}\text{CO}_2$ in air through the photosynthesis), and less so in other cases (e.g., ^{13}N -ammonia is taken up by transporters at the surface of roots). It can be a challenge to detect small amount of radioactivity in a plant when the background radioactivity in close proximity to the plants is orders of magnitude higher. These high-intensity gamma ray fluxes from background signal can produce the so-called random coincidences—false signals that are produced by chance when two unrelated gamma rays are detected by the scanner within a small pre-defined coincidence time window (typically only a few nanoseconds long). These types of coincidence events are randomly distributed in the scanner's imaging field-of-view. They not only contribute to noise in the reconstructed images but also decrease the contrast of the object to be imaged. To minimize its effect, one can use smaller amount of radioactivity for imaging experiments (which is not always an option), or shield large amount of radioactivity such that the gamma rays produced by them would not reach the detectors in the PET scanner. Figure 2 (left) shows a common bean plant being labeled by ^{13}N -ammonia through its roots. The entire root section and the ^{13}N -ammonia solution are enclosed by a lead container to shield the majority of the radioactivity from the PET scanner's detectors. Only the shoots of the plant are placed inside the PET scanner's imaging field-of-view. This setup allows us to use a large quantity of radioactivity for labeling a plant without oversaturating the scanner with high-intensity flux of gamma rays.

An alternative to shielding the unabsorbed radioactivity is to flush out (or remove) the radioactivity in the labeling chamber (or container), as will be illustrated in several examples below. Care should be taken to minimize the contamination of radioactivity on the surface of a plant. In some cases, when the “contaminated” radioactivity cannot be removed from the plant, additional shielding materials may be used to block off the gamma rays. In some extreme cases, removal of the “hot” parts from the plant may be useful as well.

2.4 Static and Dynamic Imaging Studies

A PET scanner detects and records coincidence events continuously for the entire scanning duration. The output file typically contains list-mode data (of individual coincidence events) and tag words that keep track of additional information such as time passed, event rate (for computing and correcting for random coincidences), etc. The list-mode file can be sorted to obtain the total sum of all coincidence events which are then reconstructed to form a single image volume. This is called a static image that represents the average activity distribution in the subject during the entire scan duration. This type of study is often used when the temporal dynamics of the radiotracer distribution is not critical. For example, one may use radioactive metals to evaluate the uptake of micronutrients and their allocation in plants. The kinetics may be extremely slow, and as a result, only an image acquired at a late time point is necessary.

Alternatively, the list-mode data may be broken down into multiple frames and reconstructed to form dynamic images. These image sequences represent the kinetics of radiotracer distribution. Plotting the radioactivity concentration within a particular region-of-interest (ROI) in a plant as a function of time, one will obtain the so-called time-activity curve (TAC). Using TAC of different compartments (e.g., leaf vs. stem vs. root vs. seed), one can establish compartmental models for the kinetics of a radiotracer [7]. Since the signal-to-noise ratio and image quality of a nuclear image technique (such as PET) are determined by the counting statistics of an imaging study, sufficient number of counts is essential in order to render the results meaningful. For this reason, the frame duration of the early time points of a dynamic image sequence may be short when a large quantity of radioactivity is present. As the radioactivity decays away, the frame duration needs to be increased in order to maintain sufficient counting statistics at late time points. There is not an optimal (or minimal) frame duration defined for dynamic PET imaging studies because it depends on many factors such as the object size, the pattern of activity distribution, etc. If the radioactivity is highly localized in several point- or line-like locations, they can be clearly imaged even with very short frame durations. In contrast, widely distributed radioactivity will require a longer scan duration in order to obtain sufficient counting statistics to make the data meaningful. The optimal frame duration and imaging protocol are often found experimentally. Longer frame durations and fewer numbers of frames have the benefits of fewer images to analyze and better counting statistics in each image frame. The drawback is a poor temporal resolution that may not reveal the details of a fast-changing dynamics.

Most PET scanners apply “decay correction” to dynamic PET images to account for the radioactive decay of the “signal” (i.e., the radiolabeled biomolecules) based on the half-life ($T_{1/2}$) of the radionuclide employed. The simplest way to apply the “decay correction” is to sort the list-mode data to form time frames within

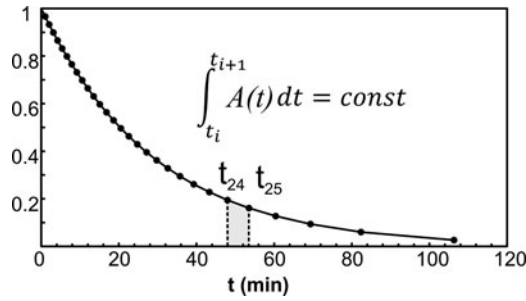


Fig. 3 Decay curve for ^{11}C ($T_{1/2} = 20.33$ min) with “one minute” equivalent time frames

which there are (on average) the same numbers of radioactive decays in each frame. That is, the duration of a time frame is increased as radioactivity decays away (*see* Fig. 3).

This approach keeps the statistical image quality the same for all frames. In many of the examples below, we use “one-minute equivalent” time frames where the initial frame was 1 min long and the subsequent frames’ duration was increased according to the decay curve defined by $T_{1/2}$. It should be noted that for every decay constant, there can be only a fixed number of such frames as the duration of the last frame may go to infinity before completion. To make use of the data from late time points, we may replace the “one-minute equivalent” frame definition by a “half-minute equivalent” or even a “quarter-minute equivalent” frame with the corresponding scaling (~ 2 and ~ 4).

“T-equivalent” time frames can be calculated using a recursive formula:

$$t_{n+1} - t_n = -\tau \times \ln \left[1 - \frac{\text{const}}{\tau} e^{t_n/\tau} \right]$$

where τ is the decay constant of the radionuclide and the *const* is the integral from 0 to time T for the initial frame:

$$\text{const} = \int_0^T e^{-t/\tau} dt$$

An example of the dynamic images is shown in Fig. 11 where a maize plant was labeled with $^{11}\text{CO}_2$ to show the kinetics of carbon translocation from a leaf to roots.

2.5 Protocols for Administering Radiotracer to Plants for PET Imaging

2.5.1 Administering Radioactive Liquid Solution Through Roots

One of the most common routes to administer radiotracers to plants is through its roots via a liquid solution containing radio-tracer. For example, ^{15}O -water [21], ^{13}N -labeled nitrate [22], or ^{22}Na -labeled salt [23] can be used to measure water flow from roots to shoots, or the uptake of N or Na by a plant’s roots, respectively. If a plant is grown hydroponically, simply mixing the radiotracer solution with the hydroponic solution would be sufficient. However, the amount of radiotracer taken up by the roots

may only be a small fraction of the total radioactivity in the hydroponic solution. As a result, it may be a challenge to quantify the radiotracer uptake in the roots when a large amount of radioactivity is still in the solution and appear as background noise in the PET images. To reduce this “noise” signal, one can drain the hydroponic solution containing radioactivity, and then refill the container with fresh hydroponic solution. This will clear the background “noise” in the PET images, simplify the data analysis, and improve the quantitative accuracy. Some radiotracer may stick to the surface of roots (or other objects) once in contact. In this case, draining the radioactive solution in the pot does not guarantee that the measured PET signals represent true uptake of the radiotracer by plants. Thus, it is important to understand the chemical property of the radiotracer before interpreting the biological meanings of PET images.

For plants grown in regular soil, one can infuse the solution containing radiotracer into the soil slowly if the half-life of the radionuclide permits such slow operation. Otherwise, pouring or injecting the solution into the pot would be sufficient. Draining the radioactive solution or flushing out radioactivity by adding more fresh water usually does not clean out the background radioactivity in the soil. Therefore, if the main interest of the study is to measure and/or to quantify the uptake of radiotracer in the roots, growing plants in hydroponic solution will be the preferred choice.

If the main interest of the study is to measure the translocation of radiotracer from roots to shoots (or other parts above soil level), a complete isolation of the roots’ volume from the rest of the plant using additional shielding (*see* Fig. 4, item 1) is recommended. Figure 2 shows a small plant being labeled and shielded using a lead pig container with a split lid. This not only reduces the radiation exposure to personnel but also eliminates potential contamination signal from the evaporated radioactivity or escaped positrons from the solution. By limiting the gamma ray flux from radioactivity in the soil (or hydroponic solution), there are also fewer random coincidences that contribute to noise in the PET images. Care should be taken not to allow the solution to be in contact with leaves or shoots above. Otherwise, it would be difficult to distinguish the contaminated radioactivity on the surface of leaves or shoots from the translocation of radiotracer taken up by the roots.

Figure 5 shows two examples where tomato plants grown hydroponically were given ^{18}F -labeled radiotracers to their roots. The container was enclosed in a lead pig (as shown in Fig. 2). Uptake and translocation of the tracers can be clearly seen in the shoots’ section.



Fig. 4 Examples of labeling chambers: (1) A lead pig container (1/2" thick) with split lid for labeled root (see also Fig. 2); (2) A pot with a gas inlet (top) and an outlet for water supply and/or drain for labeling roots with radioactive gas; (3) An opaque whole-plant gas labeling chamber with built-in LED light on top; (4) An opaque whole-plant gas labeling chamber with gas inlet and outlet; (5–7) A variety of transparent whole-plant gas labeling chambers, 4–6 inches in diameter; (8) Plant labeling chamber with fluorescent light; (9) Labeling container for small plants

2.5.2 Administering Radioactive Gas to Roots

Nitrogen fixation by plant/microbe symbiosis can be directly measured by administering ^{13}N -labeled nitrogen gas to roots followed by PET imaging [24]. Figure 6 illustrates the procedures for administering the radioactive gas and clearing the background activity around the roots in order to obtain clean PET images. The roots of a plant are sealed inside a pot with two valves (pot #2 in Fig. 4). Granulated soil (e.g., puffed clay) should be loose enough to permit free flow of gas through the pot. The shape and property of the granulated soil should be chosen to minimize the chance of radioactive gas trapping inside the soil grains. If necessary, pre-rinse the soil to fill small air pockets on the surface of granulated soil with water which can prevent radioactive gas being trapped in these small pockets during the post-labeling water fill procedure below. Clay may be used to seal the gap between the stem of a plant and the soil-protecting cover. Alternatively, an additional plastic bag around the entire plant may be used as a reservoir of the radioactive gas when the pot is filled with water later. It is recommended that the whole labeling procedure is conducted inside a fume hood certified for (small amount of) radioactive gas exhaust as a precaution in case of a leak of radioactive gas.

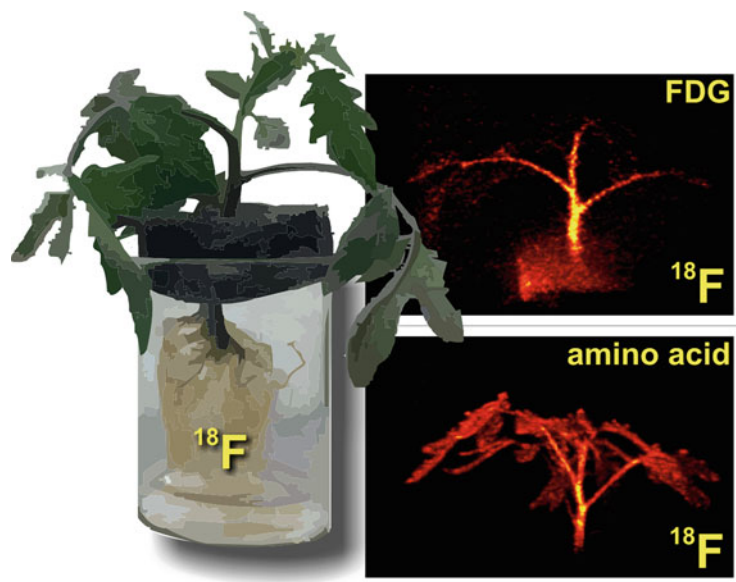


Fig. 5 Examples of radiotracer solution administration through roots: tomato plants grown hydroponically were labeled by ^{18}F FDG (top) and ^{18}F -labeled amino acids (bottom)

Radioactive gas is injected into the bottom of the pot/chamber with the top valve open for venting to balance the pressure. The vented gas may contain a small amount of radioactivity and should be kept in a sealed container rather than vented to the room. After closing both valves, wait for the uptake of radioactive gas by the plant (or plant-microbes). Upon completion of the labeling, radioactive gas is pushed out to the storage volume by pumping water into the pot to completely fill up the space in the pot. Without draining the water out, image the plant directly by a PET scanner. Water around roots is beneficial for the PET imaging since it forces the escaped positrons to annihilate near their origins.

Similar procedures can be applied to plants that are grown in hydroponic solutions or plants that are exhumed from soil and maintained in hydroponic solutions. Figure 7 shows examples where ^{13}N -labeled nitrogen gas was given to a young soybean plant that was initially grown in soil and has developed root nodules before it was transferred to hydroponic solution. The labeling chamber was sealed completely (including the small hole around the stem) and filled with hydroponic solution. ^{13}N -labeled nitrogen gas was fed to the inlet on top, while the hydroponic solution was pumped out from the chamber to suck the radioactive gas in. We waited 10 min to allow the fixation of nitrogen gas by root nodules before pumping the hydroponic solution back into the chamber to push out the radioactive gas. The entire plant and chamber (filled with hydroponic solution) was imaged by a PET scanner to obtain the images in Fig. 7 (right).

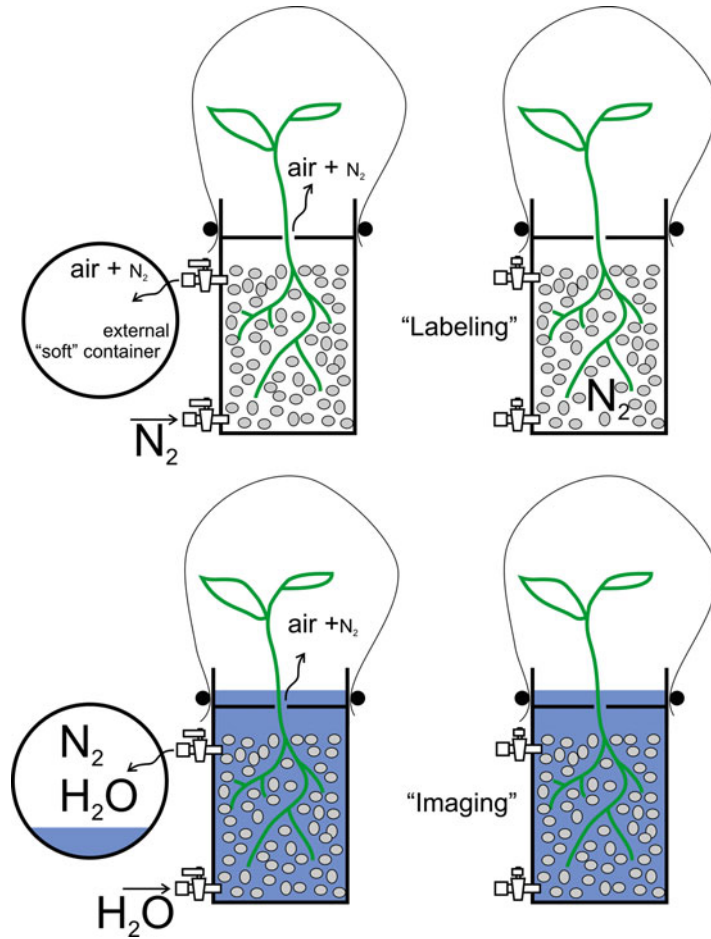


Fig. 6 Procedures for administering radioactive gas (such as ^{13}NN) to roots

2.5.3 Administering Radioactive Gas to Canopy

Using $^{11}\text{CO}_2$ as a radiolabeled tag, one can quantify carbon assimilation by plants and the efficiency of photosynthesis [25]. By tracking the ^{11}C -labeled photosynthates, one can also study the kinetics of carbon translocation and partitioning throughout a whole plant using PET imaging [8, 26]. $^{11}\text{CO}_2$ can be given to the entire plant, to a single leaf, or to a single spot of a leaf using the following procedures.

Whole-Plant Labeling

A whole plant is enclosed inside a labeling (top) chamber that is attached to a sealed pot that houses a plant (see #6 in Fig. 4). Soil can be isolated from the top labeling chamber by plastic films to minimize diffusion of the radioactive gas into the soil. In order to accommodate the entire shoots and leaves, the labeling chamber often has a large volume. As a result, clearing the background radioactive gas from the labeling chamber will take an extended period of time. Thus, this is not considered a pulse-chasing experiment [27] where a tagged tracer is administered as a single “pulse

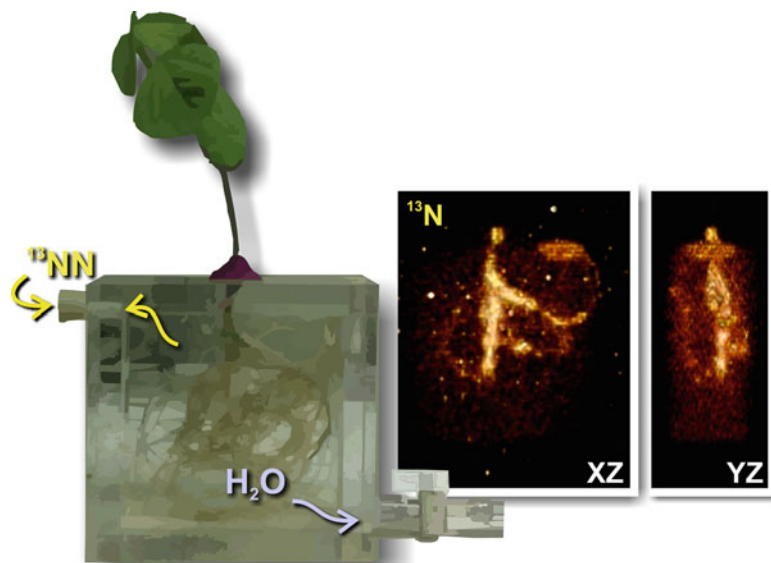


Fig. 7 Example of a soybean plant with its roots labeled with ^{13}N gas using the procedures in Fig. 6. PET images were acquired and projected onto the XZ and YZ planes for display

input” over a very short period of time and its kinetics tracked when additional “cold” (or unlabeled) tracer (such as regular photosynthate) is continuously administered to “chase” the labeled tracer. In this case, the radiotracer is assimilated into the plant continuously (with a rate that decreases over time).

Administering of radioactive gas into a labeling chamber is typically through injection using a syringe. With a syringe, one can measure precisely the amount of radioactivity in it using a dose calibrator. The highly concentrated radioactivity in a syringe can also be better shielded from radiation safety’s consideration. We typically receive up to 740 MBq (20 mCi) of $^{11}\text{CO}_2$ in a 20 mL syringe from our cyclotron facility. Using a tungsten shield to protect the operator, the radioactive gas is injected into the labeling chamber quickly by hand or slowly by an infusion pump, depending on the type of study.

In a “static” labeling, the activity is injected into the sealed labeling chamber (similar to #6 in Fig. 4). A small balloon can be connected to the outlet of the labeling chamber to prevent pressure increase that breaks the seal and causes leakage of radioactive gas. The radioactive gas remains in the labeling chamber during the imaging session. No evacuation is needed. This type of study is perhaps the simplest to carry out and is most useful if the main interest is to measure the translocation of ^{11}C -labeled photosynthates to the roots. Thus, clearing of the background $^{11}\text{CO}_2$ in the labeling chamber is less critical.

In a “dynamic” labeling, the $^{11}\text{CO}_2$ is infused into the inlet of the labeling chamber, while the air is continuously evacuated by a vacuum pump through the outlet of the chamber. The chamber is kept under a small negative pressure (relative to the atmospheric pressure). Injection can be fast via a single shot or slowly using a syringe pump. The dynamic labeling has a few advantages. Firstly, the negative pressure inside the chamber prevents the radioactivity leak outside the chamber and creates counter airflow in the pot soil (preventing radioactive gas diffusion into the soil). Secondly, the flow of the fresh air into the labeling chamber keeps the moisture, temperature, and CO_2 concentration the same as those in the plant growth chamber. Finally, the slow injection creates a quasi-stationary radioactivity concentration inside the chamber during the labeling procedure.

An example of the static whole-plant labeling (of two maize plants) is shown in Fig. 8 using pot #7 in Fig. 4. With $^{11}\text{CO}_2$ continuously being assimilated into the leaves, the radioactivity concentration in leaves increases over time to produce higher pixel values in the PET images. It should be noted that the translocation of ^{11}C -labeled photosynthates to roots takes time. With strong signals in the shoots, it is difficult to visualize the signals in the root section until late time points where the uptake in roots becomes high enough to be visualized. As mentioned before, the whole-plant labeling is not a pulse-chasing experiment, thus sub-optimal for modeling the translocation of photosynthates to roots. Single-leaf labeling below is more suitable for this type of studies.

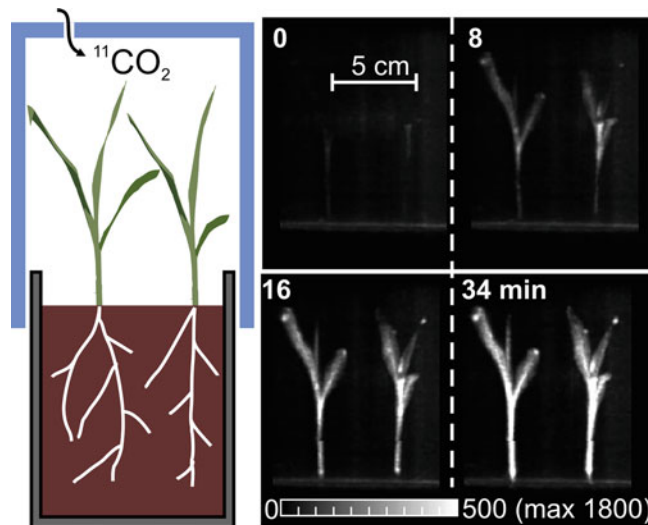


Fig. 8 Examples of PET images after whole-plant labeling: Two corn seedlings statically labeled by >740 MBq (20 mCi) of $^{11}\text{CO}_2$ injected into a labeling chamber. PET images on the right show increased uptake of ^{11}C in shoots over time

Single-Leaf Labeling

Single-leaf labeling is perhaps the most versatile procedure among all for the following reasons: (1) it requires a smaller amount of radioactivity (when compared to whole-plant labeling); (2) the labeling zone can be easily isolated (shielded) during an experiment; and (3) the labeling procedure can be short and local that is important for the dynamic studies (“pulse-chasing” techniques). We have developed multiple leaf labeling chambers (LLC) as shown in Fig. 9 with labels 1, 2, and 3. These chambers are made of two transparent acrylic plates with “open cell” neoprene gasket on each plate. A leaf to be labeled is sandwiched between those gaskets. The volume of the labeling chambers is approximately 10 mL. Since the complete sealing of the chamber is impossible without strong pressure (and potential damage) to a leaf, the chamber is maintained under a negative pressure by an external vacuum pump (through the pair of outlets that are labeled as #5 in Fig. 9) throughout the entire labeling procedure. The continuous airflow into the LLC between the semi-sealed gaskets (including the air sucked in through the “open cell” neoprene gasket itself) prevents the radioactive gas to leak out from the chamber. The gently applied imperfect seal by the neoprene gasket prevents an accidental collapse of the acrylic plates that overcompress the leaf. The continuous airflow from the outside into the labeling chamber also helps to keep the leaf under the same conditions (CO_2 concentration, moisture, temperature, etc.) as the rest of the plant in the plant growth chamber before and after labeling.

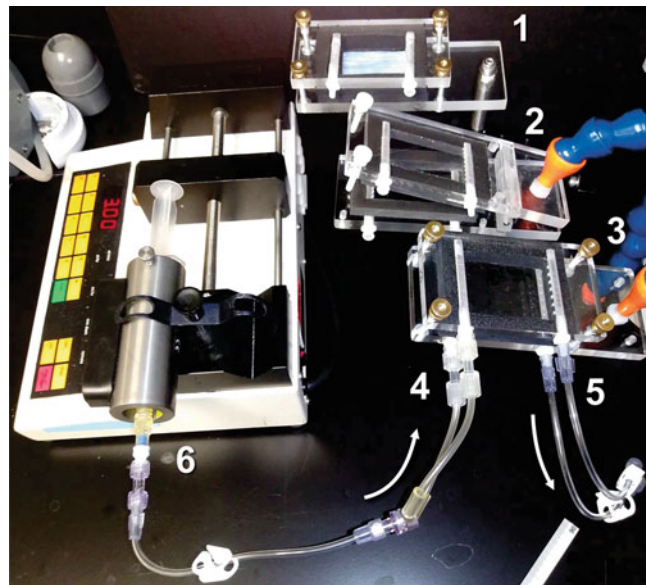


Fig. 9 Variety of chambers custom designed for single-leaf labeling: (1) vapor labeling chamber, (2) clamshell chamber, (3) split chamber. (4) and (5) are gas “inlets” and “outlets”. (6) Syringe pump with a syringe surrounded by a tungsten shield

With LLC #2 and #3, the radioactive gas is steadily injected into the LLC using a syringe pump (labeled as #6 in Fig. 9) through the two gas dispensers (set of small holes drilled through both acrylic plates) from one side of the LLC (labeled as #4 in Fig. 9) and evacuated from the other side (labeled as #5). The LLC #2 operates like a clamshell and can be opened and closed quickly for a convenient labeling operation. The LLC #3 can be completely split to allow more accurate positioning of a leaf. Its top plate can be removed without disturbing a plant during the imaging experiment. By adjusting the injection and evacuation rates, one can control a steady concentration of the labeling radioactive gas inside the LLC. Labeling time is determined by the syringe pump speed. In our experiments, we usually use 20 mL syringe (with 185 MBq (5 mCi) of $^{11}\text{CO}_2$) with 5 min injection time. Continuous evacuation and the airflow into the LLC provide relatively fast (~5 min) “wash out” of the radioactive gas from the LLC after the labeling. The radioactive gas (pumped out from the chamber) can be directed to an “activity trap” (e.g., a soda-lime trap that absorbs CO_2) or toward a storage reservoir outside of the plant growth chamber away from the PET scanner to allow radioactive decay in storage. Soda-lime traps placed into a dose calibrator can be used for real-time monitoring of radioactivity flow. Quantitative estimation of the evacuated unlabeled radioactivity is possible using a “fresh” soda-lime trap since soda-lime compound degrades quickly with the moisture from the plant growth environment. Since radioactivity may be trapped along the exhaust line (e.g., $^{11}\text{CO}_2$ dissolved in condensed water droplets on the inner surface of a tubing), it is important to monitor the tubing for the presence of radioactivity.

The LLC #1 was developed to label a leaf using the vapor of ^{13}N -ammonia solution [28]. The ^{13}N -ammonia solution can be injected by a needle through the two gaskets into a recessed pocket centered at the bottom acrylic plate. The LLC is held horizontally to prevent a spill of the radioactive solution. To increase the surface area and aid the evaporation, porous tissues (e.g., filter paper) can be placed in the pocket to absorb the radioactive solution. It is recommended to keep the chamber at slightly negative pressure to avoid accidental radioactivity leakage from the chamber.

With the single-leaf labeling, one can easily place the LLC outside of the scanner’s imaging field-of-view. The LLC can also be shielded or removed without disturbing the plant or the imaging session. This makes the single-leaf labeling protocol a convenient way for dynamic PET imaging studies of the phloem transport. With the low level of background activity in the imaging field-of-view, it becomes possible to visualize and quantify low level of radioactivity in a plant (including roots) using short time frames to analyze the kinetics of radiotracer and to model plant physiology. Figure 10a is an example of a common bean leaf being labeled with

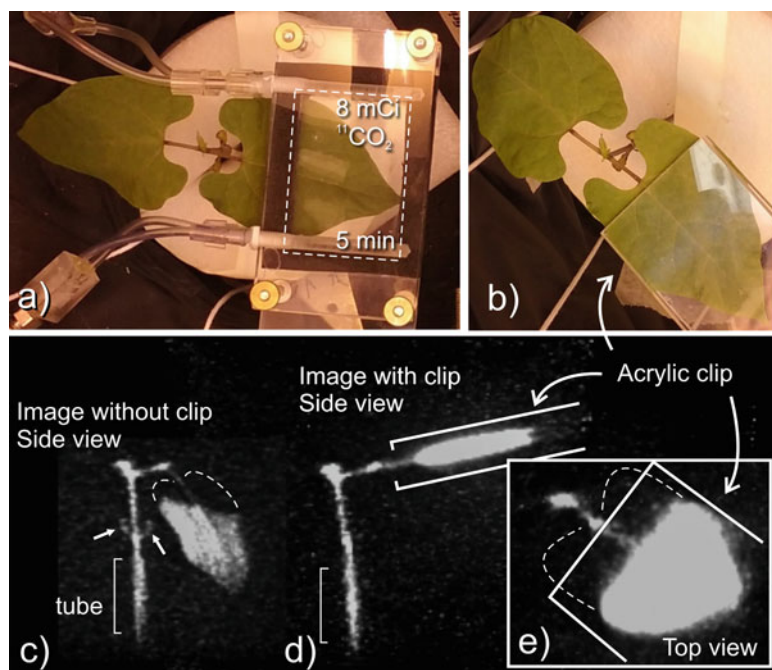


Fig. 10 (a) A common bean plant was labeled with 296 MBq (8 mCi) of $^{11}\text{CO}_2$ for 5 min using the single-leaf labeling procedures. (b) The plant was placed in a PET scanner with two acrylic plates clipped to the labeled leaf. (c) PET images of the plant acquired at 2 h 45 min post labeling. The acrylic plates were not clipped to the leaf; (d) and (e) PET images of the plant acquired at 2 h and 30 min post labeling, with the acrylic plates clipped to the leaf to force all escaped positrons to annihilate locally to improve quantitative accuracy

$^{11}\text{CO}_2$ using LLC #3. Figure 10b shows the LLC #3 completely removed after $^{11}\text{CO}_2$ labeling, and the labeled leaf sandwiched by an acrylic clip to force positrons in a leaf to annihilate for quantitative PET imaging.

Spot Labeling

The same leaf labeling chambers (LLC) above can be used to label a single spot of a long leaf blade such as those of maize plants. Figure 11 illustrates the use of dynamic PET imaging to model the translocation of photosynthates from shoots to roots in a maize plant after a single leaf was spot labeled with 185 MBq (5 mCi) of $^{11}\text{CO}_2$ for 5 min using the setup in Fig. 11a. The maximal intensity projection (MIP) images of roots (derived from 3D PET images) at different time points (relative to the injection time of the $^{11}\text{CO}_2$) are shown in Fig. 11b. We drew six cubic volumes-of-interest (VOI) in the 3D image volume (Fig. 11d) to calculate the mean radioactivity concentration in these VOI. The mean values are plotted as a function of time to obtain the “time-activity curves” (TAC) as shown in Fig. 11c. We used log-scale for the Y-axis because the activity concentration in different VOI spans a wide range that may differ by several orders of magnitude. The gray scale

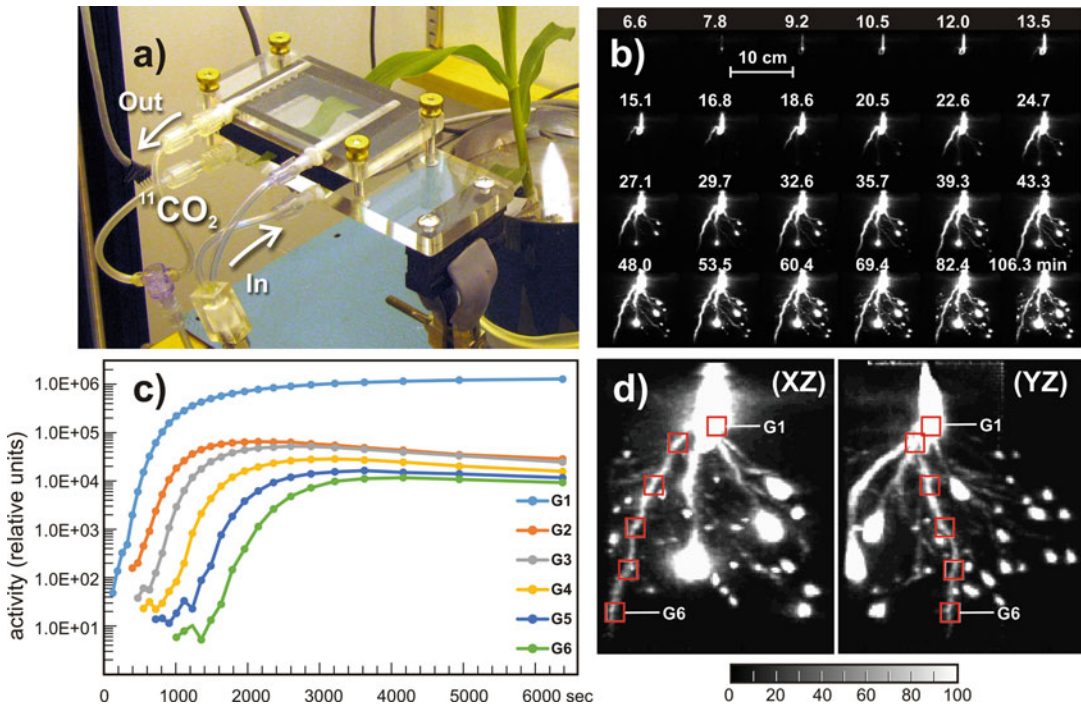


Fig. 11 Example of a pulse-chasing PET imaging experiment using “spot” labeling of a corn leaf: (a) The tip of a corn leaf was labeled with 185 MBq (5 mCi) of $^{11}\text{CO}_2$ using LLC #3 for 5 min; (b) Dynamic PET images of the roots at different time points (in min) post $^{11}\text{CO}_2$ administration; (c) Time-activity curves extracted from dynamic PET images using volume-of-interest (VOI) shown in the maximum intensity projection images (XY and YZ) in (d). Color scales used for sub-panels (b) and (d) are oversaturated to show the fine structures of the roots

of the images in Fig. 11b, d was oversaturated to reveal the minute uptakes in fine root structures. Based on the physical distances between adjacent VOIs in Fig. 11d and the difference in time between the rising-edge of the corresponding TACs in Fig. 11c, one could estimate the speed of the translocation of ^{11}C -labeled photosynthates in phloem. This type of direct measurement of physiological parameters is hard to gain using other types of imaging or analytical techniques.

3 Conclusions

PET imaging is not a common tool that plant biologists would use on daily basis for phenotyping research. However, PET offers some unique capabilities that are otherwise hard to gain using other phenotypic imaging techniques. Depending on the choice of radio-tracers employed, one may use PET to measure multiple physiological parameters or functions. Figure 12 shows an example of dual-tracer labeling and imaging to reveal the translocation of

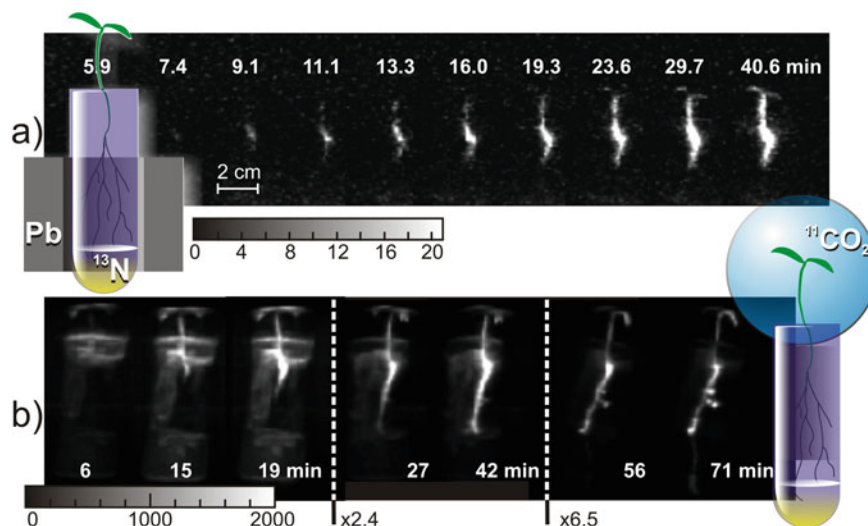


Fig. 12 Example of sequential PET imaging using different radiotracers: (a) a tomato seedling was first labeled with ^{13}N -ammonia solution administered to roots. Dynamic PET images of shoots show transport of ^{13}N through xylem; (b) 2 h later, the same plant was labeled with $^{11}\text{CO}_2$ gas administered to leaves to show ^{11}C translocation through phloem

nutrients in the xylem and phloem systems. Figure 12a shows that ~ 37 MBq (1 mCi) of ^{13}N -ammonia was administered to the roots of a small tomato seedling to measure the flow in the xylem. Only the top part of the plant was imaged by the PET scanner, while the bottom part of the plant was in a lead shield outside of the scanner's field-of-view. Two hours after the initial injection of the ^{13}N -ammonia (i.e., more than ten half-lives of the ^{13}N), the leaves and shoots of the tomato seedling were labeled by $^{11}\text{CO}_2$ to measure the ^{11}C -photosynthates flow in the phloem (Fig. 12b). This example clearly demonstrates the capability of PET for measuring different physiological functions using different radiotracers. It also highlights the potential benefits of short-lived radionuclides as one can probe a plant repetitively to study transient responses in plants as soon as the first radioactive probe decays away. It is important that one understands the physical, chemical, and biological properties of the radiotracer employed, as well as the unique strengths and limitations of PET imaging technique, in order to gain quantitative information from PET images to establish accurate biological models.

Acknowledgments

This work was supported in part by the US National Science Foundation (NSF DBI-1040498 and EPSCoR IIA 1430427/IIA 1430428), US Department of Energy (DOE-BER,

DE-SC0005157), National Academies Keck Futures Initiative, Washington University in St. Louis I-CARES program, and Malinckrodt Institute of Radiology.

References

- Li L, Zhang Q, Huang DF (2014) A review of imaging techniques for plant phenotyping. *Sensors* 14(11):20078–20111
- Jahnke S, Menzel MI, van Dusschoten D et al (2009) Combined MRI-PET dissects dynamic changes in plant structures and functions. *Plant J* 59(4):634–644
- Neilson EH, Edwards AM, Blomstedt CK et al (2015) Utilization of a high-throughput shoot imaging system to examine the dynamic phenotypic responses of a C4 cereal crop plant to nitrogen and water deficiency over time. *J Exp Bot* 66(7):1817–1832
- Sankaran S, Khot LR, Espinoza CZ et al (2015) Low-altitude, high-resolution aerial imaging systems for row and field crop phenotyping: a review. *Eur J Agron* 70:112–123
- Andrade-Sanchez P, Gore MA, Heun JT et al (2014) Development and evaluation of a field-based high-throughput phenotyping platform. *Funct Plant Biol* 41(1):68
- Chen D, Neumann K, Friedel S et al (2014) Dissecting the phenotypic components of crop plant growth and drought responses based on high-throughput image analysis. *Plant Cell* 26(12):4636–4655
- Cherry SR, Sorenson JA, Phelps ME (2003) *Physics in nuclear medicine*, 3rd edn. Saunders/Elsevier, Philadelphia
- Pickard WF, Minchin PEH, Troughton JH (1978) Real time studies of carbon-11 translocation in moonflower: I. The effects of cold blocks. *J Exp Bot* 29(4):993–1001
- Magnuson CE, Fares Y, Goeschl JD et al (1982) An integrated tracer kinetics system for studying carbon uptake and allocation in plants using continuously produced (CO₂)-C11. *Radiat Environ Biophys* 21(1):51–65
- Minchin PE, Liang A, Thorpe MR (1983) Dynamics of cold induced inhibition of phloem transport. *J Exp Bot* 34(139):156–162
- Kiser MR, Reid CD, Crowell AS et al (2008) Exploring the transport of plant metabolites using positron emitting radiotracers. *HFSP J* 2(4):189–204
- Garbout A, Munkholm L, Hansen S et al (2012) The use of PET/CT scanning technique for 3D visualization and quantification of real-time soil/plant interactions. *Plant Soil* 352(1–2):113–127
- Wang Q, Mathews AJ, Li K et al (2014) A dedicated high-resolution PET imager for plant sciences. *Phys Med Biol* 59(19):5613–5629
- Mincke J, Courtyn J, Vanhove C et al (2021) Guide to plant-PET imaging using (CO₂)-C-11. *Front Plant Sci* 12:602550
- Nakanishi H, Bughio N, Matsuhashi S et al (1999) Visualizing real time [C-¹¹]methionine translocation in Fe-sufficient and Fe-deficient barley using a Positron Emitting Tracer Imaging System (PETIS). *J Exp Bot* 50(334):637–643
- Beer S, Streun M, Hombach T et al (2010) Design and initial performance of PlanTIS: a high-resolution positron emission tomograph for plants. *Phys Med Biol* 55(3):635–646
- Weisenberger AG, Kross B, Lee SJ et al (2013) Nuclear physics detector technology applied to plant biology research. *Nucl Instrum Methods Phys Res A* 718:157–159
- Alexoff DL, Dewey SL, Vaska P et al (2011) PET imaging of thin objects: measuring the effects of positron range and partial-volume averaging in the leaf of *Nicotiana tabacum*. *Nucl Med Biol* 38(2):191–200
- Tran TM, Hampton CS, Brossard TW et al (2017) *In vivo* transport of three radioactive [F-¹⁸]-fluorinated deoxysucrose analogs by the maize sucrose transporter ZmSUT1. *Plant Physiol Biochem* 115:1–11
- Matsunami H, Arima Y, Watanabe K et al (1999) N-¹³-nitrate uptake sites and rhizobium-infectible region in a single root of common bean and soybean. *Soil Sci Plant Nutr* 45(4):955–962
- Kiyomiya S, Nakanishi H, Uchida H et al (2001) Light activates (H₂O)-O-¹⁵ flow in rice: detailed monitoring using a positron-emitting tracer imaging system (PETIS). *Physiol Plant* 113(3):359–367

22. Kawachi T, Nishijo C, Fujikake H et al (2002) Effects of anion channel blockers on xylem nitrate transport in barley seedlings. *Soil Sci Plant Nutr* 48(2):271–277
23. Arino-Estrada G, Mitchell GS, Saha P et al (2019) Imaging salt uptake dynamics in plants using PET. *Sci Rep* 9:18626
24. Ishii S, Suzui N, Ito S et al (2009) Real-time imaging of nitrogen fixation in an intact soybean plant with nodules using ^{13}N -labeled nitrogen gas. *Soil Sci Plant Nutr* 55(5): 660–666
25. Moorby J, Ebbert M, Evans NTS (1963) The translocation of the ^{11}C -labelled photosynthate in the soybean. *J Exp Bot* 14:210–220
26. Thorpe MR, Minchin PEH (1991) Continuous monitoring of fluxes of photoassimilate in leaves and whole plants. *J Exp Bot* 42(237): 461–468
27. Bacher A, Chen F, Eisenreich W (2016) Decoding biosynthetic pathways in plants by pulse-chase strategies using $(\text{CO}_2)\text{-C-}^{13}$ as a universal tracer. *Metabolites* 6(3):21
28. Li G, Lin RS, Egekwu C et al (2020) Seasonal nitrogen remobilization and the role of auxin transport in poplar trees. *J Exp Bot* 71(15): 4512–4530



Phenotyping Complex Plant Structures with a Large Format Industrial Scale High-Resolution X-Ray Tomography Instrument

Keith E. Duncan and Christopher N. Topp

Abstract

Phenotyping specific plant traits is difficult when the samples to be measured are architecturally complex. Inflorescence and root system traits are of great biological interest, but these structures present unique phenotyping challenges due to their often complicated and three-dimensional (3D) forms. We describe how a large industrial scale X-ray tomography (XRT) instrument can be used to scan architecturally complex plant structures for the goal of rapid and accurate measurement of traits that are otherwise cumbersome or not possible to capture by other means. The combination of a large imaging cabinet that can accommodate a wide range of sample size geometries and a variable microfocus reflection X-ray source allows noninvasive X-ray imaging and 3D volume generation of diverse sample types. Specific sample fixturing (mounting) and scanning conditions are presented. These techniques can be moderate to high throughput and still provide unprecedented levels of accuracy and information content in the 3D volume data they generate.

Key words X-ray tomography, Phenotyping, Excavated root system, Panicle, Inflorescence

1 Introduction

Plant phenotyping is in the midst of a technological revolution driven by enhanced sensing, imaging, and computational analysis tools [1–3]. Linking extensive genomics resources with the accurate and precise measurement of often complicated traits across cell, organ, tissue, plant, and ecological scales is the basis of the emerging field of phenomics [4]. A key goal of this work is to identify genetic mechanisms (single genes, gene families, quantitative trait loci) that control or significantly influence traits of interest and mobilize them into modern plant breeding systems for improved crop production in the face of mounting environmental challenges. Recent plant phenotyping approaches have predominately relied on optical two-dimensional (2D) image capture

technologies to provide data on traits of interest that can be measured by image analysis tools [5–7]. However, they are constrained to superficial measurements and typically can only poorly estimate three-dimensional (3D) plant traits, especially those with geometric complexity such as most shoot and root architectures.

X-ray tomography (XRT) is an effective way of capturing complex topological, geometric, and internal morphological information and providing rich 3D data sets that allow measurement and analysis of a wide range of plant traits in ways not possible with any 2D imaging platform. The first X-ray computed axial tomography (CAT) instruments appeared in the 1970s for medical imaging and typically operated in the 1 mm resolution range [8]. In the early 1980s, researchers began to develop higher-resolution instruments, “X-ray micro-computed tomography” or “X-ray microCT,” that could image in the 20 μm range for materials science [9]. Currently, XRT is referred to relative to three levels of resolution, CT, microCT, and nanoCT, which reflect the range of imaging resolutions possible across the spectrum of XRT instruments rather than referring to the specific instruments themselves [*see* Stock (2009) [8] for an excellent review of X-ray imaging technology and application].

For XRT in general, samples are placed between an X-ray generation system (source) and a device for capturing X-rays that travel through the sample (detector), today achieved with a scintillator-coated digital panel, whereas this was done with film in the past. Digital 2D images, or radiographs, are projected onto the detector as X-rays pass through the sample and are differentially absorbed due to variation in sample density (X-ray attenuation). The 3D volume is produced by collecting dozens to thousands of 2D digital radiographs as either the sample or the source-detector unit rotate over (typically) 360° . The 2D images are computationally reconstructed to generate a fully detailed 3D volume, where each 16-bit voxel (3D volume pixel) can represent 65,536 shades of gray. This grayscale information represents the physical density of various parts of the sample that form the contrast of objects of interest relative to the background, providing the means for downstream image processing and trait measurement operations, generally referred to as “feature extraction.” These methods have been employed for root phenotyping [10, 11], and a range of image analysis strategies for 3D volume data are already being tested [12–16]. Additional examples of using XRT imaging for plant phenotyping include assessment of stalk strength in maize [17–19] and sorghum [20]; grain analysis in barley [21], wheat [22], rice [23], and sorghum [24]; and measurement of developing shoots in garlic cloves [25]. Here we describe a variety of phenotyping methods that utilize XRT to generate detailed 3D volume data. The manifold downstream feature extraction pipelines are necessarily tailored to specific researchers’ application and outside the scope of this chapter.

The methods presented here deal with samples imaged in air-inflorescences and root systems that have been fully or partially excavated and washed rather than studied in situ. The root systems of many plants become woody through lignification during development, preserving well the 3D structure after excavation [10, 11, 14]. There is a growing body of research with the aim of using XRT to scan roots in various media, and to computationally separate root architecture from the surrounding growth medium (segmentation) for trait analysis in feature extraction pipelines [26–29]. Although progress with relatively small pots in low numbers has been demonstrated [30, 31], significant challenges remain for imaging and segmentation of larger root systems like maize or sorghum. Roots imaged in situ will be discussed in future works.

2 Materials

2.1 Instrumentation

There are numerous manufacturers of XRT instrumentation, including but not limited to Bruker, General Electric (GE), Nikon, North Star Imaging (NSI), Rigaku, VisiConsult, Yxlon, and ZEISS, all with useful features and capabilities. We will describe the specific instrument we are currently using for plant phenotyping; however, many of the features are universal to most commercially available systems. For example, most systems have a tiling or stitching capability whereby samples that exceed the detector dimensions at the desired resolution can be fully imaged with multiple overlapping scans that are computationally assembled in the reconstruction or post-processing steps.

We are using a model X5000 instrument from NSI (Rogers, MN) for working with a wide range of plant samples. The advantages of this industrial scale XRT instrument for our phenotyping purposes are as follows:

1. A large internal cabinet (292 cm wide by 203 cm deep by 248 cm tall) to accommodate a wide variety of sample sizes, from large whole plants down to individual plant parts.
2. An X-RAY WorX microfocus reflectance X-ray source with tungsten target capable of 10–225 kV energies.
3. A 20×24 cm Varian detector (1536×1920 pixels) with $127 \mu\text{m}$ pixel pitch capable of biological resolutions down to $\sim 25 \mu\text{m}$ voxel resolution, depending on detector-sample-source geometry and sample feature density. The detector orientation can be switched between portrait and landscape according to sample size requirements.
4. A rugged design combining durable parts and engineering with precise motion control, originally optimized for aerospace, automotive, and electronics industry applications where large and often very heavy parts are scanned.

5. A range of scanning modes: *Standard*, where the source, sample, and detector stay in one horizontal alignment during sample rotation, and the sample image remains within the borders of the detector; *Helical*, where the linked source and detector move from sample base to apex during multiple sample rotations, generating a helical scan for samples taller than the detector at the desired resolution; *Tile*, a tiling/stitching program that allows the detector to be repositioned in a grid layout of multiple overlapping 360° scans for imaging samples wider and/or taller than the detector at the desired resolution.
6. The NSI software package generates full 3D volume reconstructions and provides data export options across a range of file formats for use in feature extraction pipelines.

2.2 Supplies

1. Extruded/expanded polystyrene (EPS). Widely available in light blue or pink sheets of varying thickness from large hardware and supply stores in North America. This material is functionally transparent to most X-ray energies, and the samples mounted with EPS fixtures can be imaged as though “floating in air.” This is important as metal fixtures contained within the scan can produce imaging artifacts in the reconstructed volume, particularly with low density samples.
2. Portable clamping system. There are numerous examples of portable clamps or vises on the market; useful features include a wide adjustable grip range, rubber or similar coated clamps, a swivel mount to adjust clamped sample angle, and a weighted base for stability. We use portable clamping units from Pana-Vise (Reno, NV).
3. Carpet anti-slip fabric. This tacky rubberized material typically comes in rolls for easy cutting with scissors, is available from hardware and household goods stores, and is used underneath carpets to prevent slipping. Squares of anti-slip fabric can be used to prevent fixtures from slipping on the turntable when the sample is rotating during scans.
4. Lab jack. These articulated platforms are useful for raising and lowering fixtures to keep samples optimally positioned between the source and detector, especially for helical scans. By placing squares of EPS with a piece of anti-slip fabric on top of the lab jack, samples can be raised into the X-ray beam without any visible metal structures projected onto the detector during scanning. Anti-slip fabric will also keep the lab jack from slipping on the metal turntable during rotation.
5. Utility knife. Used for carving EPS blocks to secure and wedge samples into fixtures; ideally choose a retractable box cutter style to reduce injury risk.
6. Drywall or jab saw. Handsaw for cutting large blocks of EPS into useful sizes and shapes.

7. Gaffer's tape. Gaffer's tape is typically the black cloth style, tears easily across and along its length, and does not leave adhesive residue like silver duct tape. This tape will be visible in most low energy scanning protocols so effort should be made to keep it on the outside of the sample region of interest so it can be easily cropped out of the final scan volume.
8. Hot glue gun. A conventional hobby shop or hardware store hot glue gun is useful for assembling fixtures or making stacks of EPS that can be carved to support samples of unusual geometry. Use the glue on the fixtures and EPS shapes, and avoid using the glue on the sample itself.

3 Methods

This chapter is focused on XRT scanning methods and parameters; therefore, specific details on growth chamber, greenhouse, or field conditions used to grow and harvest the plants are left to the requirements of individual researchers.

3.1 Fixturing

It is useful for all plant samples to have some length of stalk, stem, petiole, or other structure to serve as a point of attachment for mounting inside the XRT instrument. Larger samples like excavated maize and sorghum root crowns (*see Note 1*) or mature sorghum and quinoa panicles (*see Note 3*) can be clamped directly by the stalk (Fig. 1a–c). Samples with small, thin stems like grape rachises or *Setaria* panicles (*see Note 2*) can be pressed between two rectangles of EPS, and the EPS clamped with a portable vise so the panicle or other structures of interest can be raised into the X-ray beam during scanning (Fig. 1d, e). Very large root systems like excavated switchgrass (*see Note 5*) can be fixtured in a frame cut from a single large sheet of 8-cm-thick EPS, secured with a wide base, and mounted directly to the sample turntable to prevent wobble or tipping during the scan (Fig. 1f). When a convenient clamping structure is not available on the sample itself, e.g., a mature maize ear with the stalk snapped off (*see Note 4*), blocks of EPS can be layered to form a base, a hole carved in the center, and the sample wedged directly into the base (Fig. 2a). This fixture can be placed on a lab jack and raised into the X-ray beam so that just the ear image is projected onto the detector with multiple EPS layers—functionally transparent to X-ray—separating the ear from the dense metal base of the lab jack (Fig. 2b).

3.2 Scan Parameters

In general, the greater the sample density, the higher kV levels must be used to penetrate the sample. For our phenotyping work, samples are typically of low density so relatively low energies in the 50–80 kV range can be used. The goal is to use the lowest kV level necessary to see all the way through the sample, and adjust the μA



Fig. 1 Various fixtures for securely holding samples during rotation while scanning; the detector in each image is 20×24 cm for scale. **(a–c)** A portable vise is used to directly clamp a maize root crown **(a)**, and panicles of sorghum **(b)** and quinoa **(c)** in front of the detector for scanning. **(d, e)** Rectangular pieces of expanded polystyrene (EPS) are used to hold thinner samples like grape rachis **(d)** and *Setaria* panicles **(e)** in the clamp during scanning. **(f)** An excavated switchgrass root system is secured within a frame cut from a single EPS board and mounted directly to the instrument turntable

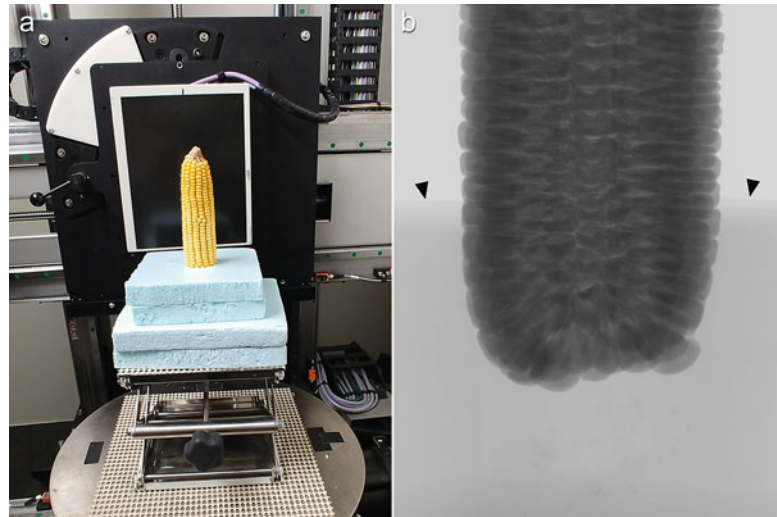


Fig. 2 (a) Mature maize ear secured within a stack of EPS blocks and mounted on a lab jack for scanning. (b) Digital radiograph from scan of a mature maize ear within blocks of EPS; note that although the maize ear is positioned within the EPS stack, EPS is functionally transparent to X-rays (arrowheads)

to control the brightness of the image on the detector. Examples of 2D digital radiographs from different plant samples are shown in Fig. 3. The size of the sample will dictate how to arrange the geometry inside the instrument so that the projected image is fully contained within the edges of the detector. Alternating the detector between landscape and portrait orientations will optimize scan geometry for varying sample sizes. For tall and narrow samples, the helical scan mode will steadily raise the linked source and detector, while the sample continuously rotates, until the entire sample length is scanned bottom to top. This is a useful compromise for imaging very tall samples (up to 1 m) that would otherwise require the stitching of numerous standard scans and significantly increase both computational difficulty and final file size. For wide and tall samples, the tiling function can be used to collect multiple overlapping 360° rotational scans in 1×2 or 2×2 tiles which the NSI software automatically stitches together into a single 3D volume afterward during reconstruction. See Table 1 for detailed scan parameters.

3.3 Data Export Options

Most commercial XRT systems use their own set of geometry and calibration tools for background adjustment of each scan to facilitate accurate measurement of reconstructed volumes. The NSI software incorporates a geometry scan into the reconstruction process and generates a 3D volume (Fig. 4) that can be exported in a variety of formats, depending on what is required for individual feature extraction pipelines. One export option is 2D slices in all

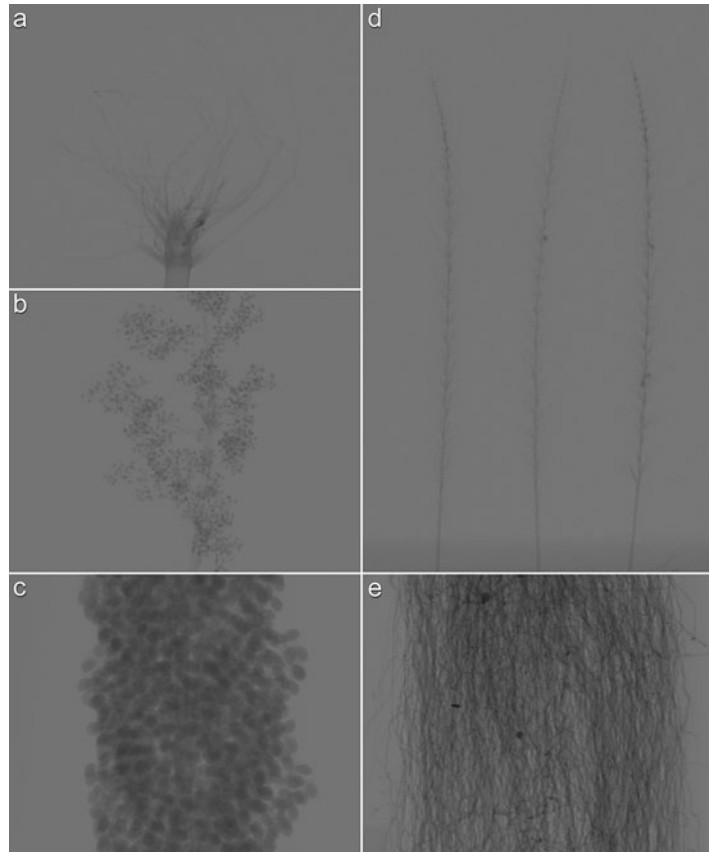


Fig. 3 Examples of 2D digital radiographs from a variety of plant samples; standard scans (**a**, **d**) where the sample projection fits entirely within the detector borders, and helical scans of tall samples (**b**, **c**, **e**) where the linked source and detector slowly travel from sample base to apex during continuous sample rotation. (**a**) Excavated maize root crown, (**b**) quinoa panicle, (**c**) sorghum panicle, (**d**) *Setaria* panicles, and (**e**) excavated switchgrass root system

three imaging planes and in a range of formats (TIFF, JPEG, BMP, PNG, DICONDE, RAW). There is the option to export a 3D volume in RAW format, which also generates a DAT file with volume size parameters. One can also use NSI software to render the 3D volume as a surface file and export over a range of formats (STL, PTY, OBJ, VRML, ASCII, DXF). With any export function, care should be taken to understand all operations performed on the original data, for example, downsampling from 16-bit to 8-bit, converting or compressing to lossy formats such as JPEG, or various options for rendering 3D volumes as surfaces.

Table 1

Summary of X-ray tomography scan parameters across a range of plant samples, using a model X5000 instrument from North Star Imaging (Rogers, MN)

Scan parameters Mode	Maize/sorghum root crowns		Grape rachis	<i>Setaria</i> panicle	Sorghum panicle	Quinoa panicle	Switchgrass root system	Mature maize ear
	Standard	Tile	Standard	Standard	Helical	Standard	Helical	Helical
kV	70	70	60	60	60	70	70	60
μA	1700	1700	1200	600	750	1700	1900	1000
fps ^a	10	10	10	10	12.5	10	10	10
Source-sample (mm)	1036	1024	767	300	400	885	943	258
Source-detector (mm)	1210	1203	912	523	625	1189	1209	705
Magnification	1.17×	1.17×	1.19×	1.74×	1.56×	1.34×	1.28×	2.73×
Voxel (μm ³)	109	108	108	73	82	95	99	46
# Projections	1800	1800	1200	5000	7500	8400	36,000	4800
Time (min)	3	6	2	8	10	28	61	131
Size ^b (GB)	17	48	14	29	50	60	96	71

^aFrames per second

^bScan files plus reconstructed 3D volume

4 Notes

Below are observations and suggestions for a range of specific plant sample types. For all these examples, tissues are most effectively scanned when they are dried. Not only are they typically lighter in weight and easier to fixture, but green tissue can also desiccate during the scan and cause blurring of the 3D volume. Also, researchers must be prepared to run pilot scans to determine the minimum parameters (energy, resolution, number of projections, scan time, file size, export format) that will still provide the required 3D volume data that yields biologically relevant features. This will vary significantly between sample types, and the examples presented here can be used as a range-finder depending on individual researchers' sample requirements, and the specific XRT instrument being used. Finally, if sample density measurements are required from the XRT scan data, density standards need to be incorporated into each scanning session to allow computational calibration of gray scale values to actual density. Commercial density standards are available from instrument manufacturers. *See* Fig. 1 for examples of sample fixturing.

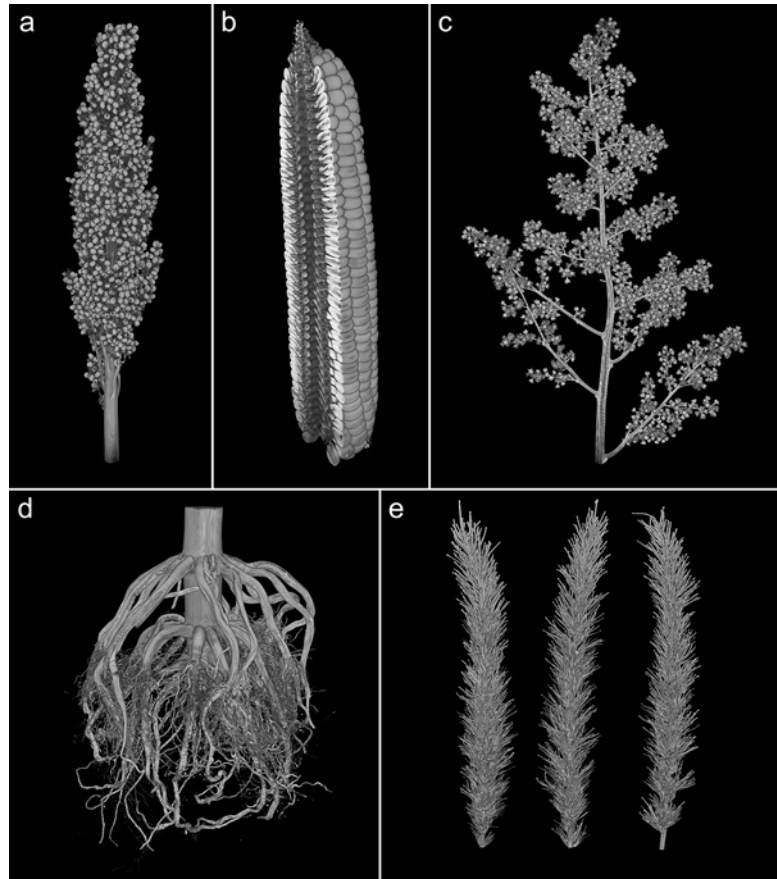


Fig. 4 Full 3D reconstructions from scans of a variety of plant samples, each one assembled from thousands of 2D digital radiographs. (a) Sorghum panicle, (b) mature maize ear (cutaway shows internal detail), (c) quinoa panicle, (d) excavated maize root crown, and (e) *Setaria* panicles

1. Excavated maize (*Zea mays* L.) and sorghum (*Sorghum bicolor*) root crowns. This approach works well with any root systems that have some level of lignification which largely preserves the 3D root crown architecture. Maize and sorghum growth conditions will vary based on the individual requirements of specific research programs, and potential methods for collecting excavated root crowns have been described [6, 10]. It is important to remove as much potting mix or field soil from root crowns as possible since any foreign material clinging to roots will be included in the scan volume, making the identification and measurement of root traits more difficult during feature extraction. When cutting off aboveground stalk tissue, be sure to leave a minimum of 15–20 cm of stalk so the sample can be clamped securely in the portable vise without wobbling or slipping during the scan. When setting up scanning geometry,

the detector must be moved as far as possible away from the source, with the sample very close to the detector without touching during rotation, to allow as much of the root crown to be projected within the borders of the detector without showing the clamps in the scan volume. This also allows for a level of scan-to-scan consistency, for example, always having 2 cm of stalk tissue above the root crown in each scan for comparing all scans across a population. If significant features of the root crown fall outside the edge of the detector, then helical or tiled/stitched scans must be used to image the entire sample. Note that these specialized scanning modes can increase final file sizes by a significant amount (*see* Table 1). Researchers should explore whether sampling at earlier time points or developmental stages might provide smaller root crowns that fit entirely on the detector but still yield biologically relevant root architecture features.

2. Grape (*Vitis* spp.) rachis. The rachis is the framework of stem tissues that bears the individual berries, and rachis branching architecture is highly variable across *Vitis* spp. [12, 13, 16]. Individual dried rachis samples can be secured between two rectangles of EPS and clamped so that the entire rachis lies within the detector frame for imaging, with enough EPS to keep the clamps out of the scan volume.
3. Panicles from sorghum, quinoa (*Chenopodium quinoa*), and *Setaria viridis*. Imaging sorghum or quinoa inflorescence structures, e.g., panicles, typically requires a helical scan as the projected panicle image is usually taller than the detector. Panicles can vary in width so the detector will have to be adjusted to either landscape or portrait orientations depending on the required resolution of the final 3D volume. The researcher must decide what features are to be measured in the scan and adjust imaging parameters accordingly. If fine details of individual grains are desired, then higher kV (greater penetration) and higher scan resolution parameters (shorter source-sample-detector distance) are required. These conditions can be used for a wide range of inflorescence structures, particularly grasses like maize, wheat, rice, barley, oats, and others, to assess a variety of traits including grain fill, seed number and distribution, and underlying panicle architecture. *Setaria* panicles are much smaller allowing multiple samples to be fixtured and scanned simultaneously, and the individual 3D volumes separated computationally before feature extraction. *Setaria* samples can be scanned by mounting one or more panicles between rectangles of EPS and clamped as described for the grape rachis above.

4. Mature maize ears. Stacks of EPS can be combined and secured with hot glue, and a hole carved in the center that supports a single maize ear. Ears should be secured by a tight fit or by using smaller pieces of EPS as wedges or shims to prevent any movement or wobble from the ear during scanning. Blocks of EPS with secured maize ears can be placed on lab jacks and raised into the X-ray beam to image the entire ear using a helical scan mode. Higher kV levels are required to adequately penetrate the ear and provide sufficient structural information about kernel number, size, shape, density, and distribution as well as relevant cob data. Again, if density information is required, a density standard should be incorporated into the scanning protocol.
5. Switchgrass (*Panicum virgatum*) excavated root systems. Excavated switchgrass root systems are large and complex, often many meters deep in situ. Removal of soil or potting mix is crucial for both reducing the weight of the root system and improving the quality of scan data. The vertical travel of the NSI X5000 has a limit of just over 1 m given the unique fixturing requirements of stabilizing such a tall sample. To capture sufficient imaging data for this sample, many thousands of images must be collected over the course of the helical scan which generates accordingly large data sets (*see* Table 1). Large fixtures of EPS are required to suspend and stabilize excavated switchgrass root systems, and the fixture should also help confine the roots to a width narrow enough to be captured by the detector in landscape mode. Ideally, use a single large piece of EPS at least 8–10 cm thick and over 1 m long; this will improve the fixture's structural strength compared to smaller pieces joined together with glue or tape, which will be visible in the 3D reconstruction. Mount the frame in an EPS base wide enough to handle the upright frame and not wobble, but not exceed the width of the turntable and potentially interfere with the detector during sample rotation. The base can be secured directly to the turntable with gaffer's tape.

Acknowledgments

This work was supported by collaborative agreements with Valent BioSciences and Sumitomo Chemical Company, and from National Science Foundation projects IIA-1355406, IOS-1638507, and DBI-1759796.

References

1. Fiorani F, Schurr U (2013) Future scenarios for plant phenotyping. *Annu Rev Plant Biol* 64: 267–291. <https://doi.org/10.1146/annurev-arplant-050312-120137>
2. Fahlgren N, Gehan MA, Baxter I (2015) Lights, camera, action: high-throughput plant phenotyping is ready for a close-up. *Curr Opin Plant Biol* 24:93–99. <https://doi.org/10.1016/j.pbi.2015.02.006>
3. Gehan MA, Kellogg EA (2017) High throughput phenotyping. *Am J Bot* 104:505–508. <https://doi.org/10.3732/ajb.1700044>
4. Houle D, Govindaraju D, Omholt S (2010) Phenomics: the next challenge. *Nat Rev Genet* 11:855–866. <https://doi.org/10.1038/nrg2897>
5. Das A, Schneider H, Burrige J et al (2015) Digital imaging of root traits (DIRT): a high throughput computing and collaboration platform for field based root phenomics. *Plant Methods* 11:51. <https://doi.org/10.1186/s13007-015-0093-3>
6. Agnew E, Bray A, Floro E et al (2017) Whole-plant manual and image-based phenotyping in controlled environments. *Curr Protoc Plant Biol* 2:1–21. <https://doi.org/10.1002/cppb.20044>
7. Gibbs J, Pound M, French A et al (2018) Plant phenotyping: an active vision cell for three-dimensional plant shoot reconstruction. *Plant Physiol* 178:524–534. <https://doi.org/10.1104/pp.18.00664>
8. Stock S (2009) *MicroComputed tomography: methodology and applications*. CRC Press, Boca Raton
9. Elliott J, Dover S (1982) X-ray microtomography. *J Microsc* 126:211–213
10. Topp C, Bray A, Ellis N, Liu Z (2016) How can we harness quantitative genetic variation in crop root systems for agricultural improvement? *J Integr Plant Biol* 58:213–225. <https://doi.org/10.1111/jipb.12470>
11. Shao M-R, Jiang N, Li M et al (2021) Complementary phenotyping of maize root system architecture by root pulling force and X-ray imaging. *Plant Phenom* 2021:2021. <https://doi.org/10.34133/2021/9859254>
12. Li M, Duncan K, Topp C, Chitwood D (2017) Persistent homology and the branching topologies of plants. *Am J Bot* 104:349–353. <https://doi.org/10.3732/ajb1700046>
13. Li M, Frank M, Coneva V et al (2018) The persistent homology mathematical framework provides enhanced genotype-to-phenotype associations for plant morphology. *Plant Physiol* 177:1382–1395. <https://doi.org/10.1104/pp.18.00104>
14. Bray A, Topp C (2018) The quantitative genetic control of root architecture in maize. *Plant Cell Physiol* 59:1919–1930. <https://doi.org/10.1093/pcp/pcy141>
15. Jiang N, Floro E, Bray A et al (2019) Three-dimensional time-lapse analysis reveals multi-scale relationships in maize root systems with contrasting architectures. *Plant Cell* 31: 1708–1722. <https://doi.org/10.1105/tpc.19.00015>
16. Li M, Klein L, Duncan K et al (2019) Characterizing 3D inflorescence architecture in grapevine using X-ray imaging and advanced morphometrics: implications for understanding cluster density. *J Exp Bot* 70:6261–6276. <https://doi.org/10.1093/jxb/erz394>
17. Du J, Zhang Y, Gao X et al (2016) Micron-scale phenotyping quantification and three-dimensional microstructure reconstruction of vascular bundles within maize stalks based on micro-CT scanning. *Funct Plant Biol* 44: 10–22. <https://doi.org/10.1071/fp16117>
18. Robertson D, Julius M, Lee S-Y et al (2017) Maize stalk lodging: morphological determinants of stalk strength. *Crop Sci* 57:926–934. <https://doi.org/10.2135/cropsci2016.07.0569>
19. Zhang Y, Du J, Wang J et al (2018) High-throughput micro-phenotyping measurements applied to assess stalk lodging in maize (*Zea mays* L.). *Biol Res* 51:40. <https://doi.org/10.1186/s40659-018-0190-7>
20. Gomez F, Carvalho G Jr, Shi F et al (2018) High throughput phenotyping of morpho-anatomical stem properties using X-ray computed tomography in sorghum. *Plant Methods* 14:59. <https://doi.org/10.1186/s13007-018-0326-3>
21. Tracy S, Gómez J, Sturrock C et al (2017) Non-destructive determination of floral staging in cereals using X-ray micro computed tomography (μ CT). *Plant Methods* 13:9. <https://doi.org/10.1186/s13007-017-0162-x>
22. Hughes N, Askew K, Scotson C et al (2017) Non-destructive, high-content analysis of wheat grain traits using X-ray micro computed tomography. *Plant Methods* 13:76. <https://doi.org/10.1186/s13007-017-0229-8>
23. Jhala V, Thaker V (2015) X-ray computed tomography to study rice (*Oryza sativa* L.) panicle development. *J Exp Bot* 66: 6819–6825. <https://doi.org/10.1093/jxb/erv387>

24. Li M, Shao M-R, Zeng D et al (2020) Comprehensive 3D phenotyping reveals continuous morphological variation across genetically diverse sorghum inflorescences. *New Phytol* 226:1873–1885. <https://doi.org/10.1111/nph.16533>
25. Comparini D, Kihara T, Kawano T (2016) Uses of X-ray 3D-computed-tomography to monitor the development of garlic shooting inside the intact cloves. *Environ Control Biol* 54:39–44
26. Mairhofer S, Pridmore T, Johnson J et al (2017) X-ray computed tomography of crop plant root systems grown in soil. *Curr Protoc Plant Biol* 2:270–286. <https://doi.org/10.1002/cppb.20049>
27. Morris E, Griffiths M, Golebiowska A et al (2017) Shaping 3D root system architecture. *Curr Biol* 27:R919–R930. <https://doi.org/10.1016/j.cub.2017.06.043>
28. Mairhofer S, Zappala S, Tracy S et al (2013) Recovering complete plant root system architectures from soil via X-ray μ -computed tomography. *Plant Methods* 9:8. <https://doi.org/10.1186/1746-4811-9-8>
29. Mooney S, Pridmore T, Helliwell J et al (2012) Developing X-ray computed tomography to non-invasively image 3-D root systems architecture in soil. *Plant Soil* 352:1–22. <https://doi.org/10.1007/s11104-011-1039-9>
30. Fry E, Evans A, Sturrock C et al (2018) Root architecture governs plasticity in response to drought. *Plant Soil* 433:189. <https://doi.org/10.1007/s11104-018-3824-1>
31. Burr-Hersey J, Mooney S, Bengough A et al (2017) Developmental morphology of cover crop species exhibit contrasting behavior to changes in soil bulk density, revealed by X-ray computed tomography. *PLoS One* 12: e0181872

Part IV

HTP Protocols for Plants Growing Under Field Conditions



Chapter 13

Challenges for a Massive Implementation of Phenomics in Plant Breeding Programs

Gustavo A. Lobos, Félix Estrada, Alejandro del Pozo, Sebastián Romero-Bravo, Cesar A. Astudillo, and Freddy Mora-Poblete

Abstract

Due to climate change and expected food shortage in the coming decades, not only will it be necessary to develop cultivars with greater tolerance to environmental stress, but it is also imperative to reduce breeding cycle time. In addition to yield evaluation, plant breeders resort to many sensory assessments and some others of intermediate complexity. However, to develop cultivars better adapted to current/future constraints, it is necessary to incorporate a new set of traits, such as morphophysiological and physicochemical attributes, information relevant to the successful selection of genotypes or parents. Unfortunately, because of the large number of genotypes to be screened, measurements with conventional equipment are unfeasible, especially under field conditions. High-throughput plant phenotyping (HTPP) facilitates collecting a significant amount of data quickly; however, it is necessary to transform all this information (e.g., plant reflectance) into helpful descriptors to the breeder. To the extent that a holistic characterization of the plant (phenomics) is performed in challenging environments, it will be possible to select the best genotypes (forward phenomics) objectively but also understand why the said individual differs from the rest (reverse phenomics). Unfortunately, several elements had prevented phenomics from developing as desired. Consequently, a new set of prediction/validation methodologies, seasonal ambient information, and the fusion of data matrices (e.g., genotypic and phenotypic information) need to be incorporated into the modeling. In this sense, for the massive implementation of phenomics in plant breeding, it will be essential to count an interdisciplinary team that responds to the urgent need to release material with greater capacity to tolerate environmental stress. Therefore, breeding programs should (i) be more efficient (e.g., early discarding of unsuitable material), (ii) have shorter breeding cycles (fewer crosses to achieve the desired cultivar), and (iii) be more productive, increasing the probability of success at the end of the breeding process (percentage of cultivars released to the number of initial crosses).

Key words Phenomics, Breeding, HTPP

1 Introduction

There is global concern about the significant gap between projected world food consumption in the coming decades and expected crop yields for those periods [1]. In the specific case of wheat, due to the

increase in world population, current production will likely have to double by mid-century [2, 3]. To achieve that production, wheat yields must increase at an annual rate of 1.6% [4], which is a far cry from the 1.26% achieved in recent years [5]. Moreover, in many agricultural areas, the current effects of global warming on annual weather patterns and atypical climatic events threaten to increase further yields [6–10]. Indeed, climate change is expected to negatively affect the area available for crop production [11, 12]. It is suggested that the most efficient way to address the food crisis is through agronomic management and genetic improvement [13]; the former has a more immediate but short-range impact due to the effects of climate change [14], while the latter requires a greater effort over time, but addresses the problem more directly [15].

The use of high-throughput plant phenotyping (HTPP) allows an indirect estimation of desirable traits through rapid, low-cost, non-destructive, and simultaneous (sometimes) assessments of thousands of genotypes. The phenotyping techniques commonly used in plant breeding programs are spectrometry and thermography [16].

In the case of spectrometry, the equipment can record the reflectance of plants and their spectral signature (i.e., graphic characterization of reflectance) related to specific traits or plant conditions [17]. On the other hand, thermography uses radiometric information to estimate plant temperature. This technology is a valuable instrument to study the spatial and temporal heterogeneity of plant water status, the photosynthetic apparatus, and the interaction with the ambient in terms of cooling capacity [18].

This book covers different HTPP tools, strategies, and methodologies. The focus of this chapter is to review the aspects that are probably limiting the massive implementation of phenomics in plant breeding programs, especially in the field. Special emphasis is placed on spectral reflectance.

2 The Relevance of High-Throughput Plant Physiology

Since the inception of civilizations, the domestication of crops has depended on man's ability to identify the individuals who best behave in each location [19]. From then on, phenotyping and selection have been the basis for plant improvement.

Due to the effects of climate change and food shortages expected in the coming decades, along with the development of cultivars with greater tolerance to environmental stress, it is also imperative to shorten the breeding cycle time (from first cross to cultivar release). The problem is that to develop tolerant cultivars, breeders must anticipate an uncertain environmental scenario. A couple of decades will probably have passed by the time the cross that gave rise to a given cultivar is massively sown/planted. In other

words, to effectively overcome environmental constraints in another 15–20 years, crossbreeding must be strategically planned today, not tomorrow.

Therefore, to effectively develop cultivars well adapted to adverse conditions, breeders will have to rely on a new set of morphophysiological and physicochemical traits, which are not always easy to implement [15, 20]; many of these traits have been related to heat and drought tolerance in Mediterranean crops [21–26]. Unfortunately, due to the high number of genotypes that must be screened (time consumption) and the complexity of the physiological measurements, traditional breeding has not fully addressed the challenge of a deep phenotype characterization [16].

This situation is more complex in breeding programs oriented to ambient stress (e.g., heat or water deficit) where, in order to evaluate individuals at their maximum expression of the response to the stimulus, the measurement window is limited to a few hours per day (± 2 h from zenith). For example, considering a set of ecophysiological traits (e.g., gas exchange, chlorophyll a fluorescence, leaf/stem water potential, and pigment content), it would be possible to phenotype a maximum of 15–20 genotypes per day [27]. Therefore, the massive use of these time-consuming assessments is impractical to consider in plant breeding, especially under field conditions. Consequently, breeders have continued with traditional selection methodologies, focusing on sensory descriptions (e.g., leaf color, resistance to diseases, growth habit, phenology) supported by information from relevant evaluations (e.g., yield and its agronomic components, grain or fruit quality, postharvest life).

In the case of marker-assisted selection (MAS), despite the technological advances in gene sequencing (i.e., routine, cheap, and fast) [28–30], for a correct analysis and interpretation of such information, it is also necessary to count toward a more complex phenotypic characterization, with the drawbacks above [31]. Therefore, MAS may not have the expected impact on plant breeding programs at the field level, where there is limited control of ambient factors. Insufficient phenotype characterization has led to the development of models with a narrow range of spatial and temporal scales (basically at the molecule level), limiting the ability to predict when the environmental scenario vary. Proof of this would be that, despite the development of MAS during the last decades, the relative gain in grain yield continues with negative growth rates [32].

Probably, at present, the best alternative for a more detailed characterization of the individual is the acquisition of high-dimensional phenotypic data (high-throughput phenotyping), which allows the researcher to generate information for a holistic understanding of plant responses, or “phenomics” [33]. Several traits are evaluated to understand the relationship between them and then describe or at least hypothesize the tolerance mechanism

(s) involved in a given genotype. In this sense, it is possible to estimate complex traits that might be relevant for genotype selection (*forward phenomics*) and understand why a given individual excels in a specific environment (*reverse phenomics*) [15].

Although, in the last decade, some researchers have proposed a physiologically oriented approach to breeding [34–36], it was not until a few years ago that the term “high-throughput plant physiology” started to be used in phenotyping and phenomics for plant breeding [31]. Compared to the traditional selection, some studies show that using physiological traits in wheat breeding programs could double the number of advanced lines of high performance under water deficit [36]. In Chile, spring wheat yield progress between 1920 and 2008 was positively related to stomatal conductance to water vapor (g_s), traits that breeders did not attempt to characterize prior to the release of each cultivar [37].

Among the many techniques to characterize physiological traits of plants growing under adverse conditions, chlorophyll a fluorescence has proven to be a powerful tool to consider in plant breeding [25, 38]. Its relevance lies in the sensitivity of this technique to distinguish differences in the photosynthetic capacity of plants subjected to abiotic stress [39, 40], identifying changes in photosynthesis’s biochemical processes (e.g., damage of photosystems or the electron transport chain), which directly limit CO_2 assimilation [41, 42]. Another relevant advantage is the time taken for each measurement (<6 s) compared to those performed with an infrared gas analyzer (<6 min) to characterize leaf gas exchange (net CO_2 assimilation rate (A_n), g_s , internal CO_2 concentration, and transpiration rate) [43–45], facilitating the performance of a more significant number of readings in a day. For example, the maximum photochemical quantum yield (F_v/F_m) of photosystem II (FSII), a parameter related to A_n [46], has been used to identify spring wheat genotypes that were tolerant to high temperatures [47, 48]. To select wild wheat genotypes tolerant to water deficit, along with F_v/F_m , other authors have also suggested reading the minimum and maximum fluorescence under dark conditions (F_0 and F_m , respectively) [49]. Working with 25 L potted blueberries subjected to heat, drought, and their combination, other fluorescence parameters were identified that were related to water deficit tolerance: maximum electron transport rate (ETR_{max}), PAR at which PSII saturates (IK), and photochemical quenching coefficients (qP and qL ; puddle and lake model, respectively) [25]. The same authors indicate that to identify the signals linked to heat stress and that of its combination with water deficit, the quantum yield of non-photochemical energy conversion in PSII due to the downregulation of the light-harvesting function [$Y(NPQ)$] and the quantum yield of non-photochemical energy conversion in PSII other than that caused by the downregulation of the light-harvesting function [$Y(NO)$] are highlighted.

Remote sensing techniques are non-destructive, rapid, and large-scale integrated (i.e., at canopy level) assessments of the crop performance. The spectrometers are capable of measuring over a wide range of the electromagnetic spectrum (i.e., ~300–2500 nm) and have been proposed as tools for crop improvement, such as in durum wheat [50]. In general, the reflectance pattern has been related to physiological and agronomic traits through different spectral reflectance indices (i.e., relationships between some wavelengths; SRI). The first SRI, known as the simple ratio (SR), included band wavelength in the visible (VIS) and near-infrared (NIR): $SR = R_{NIR}/R_{VIS}$. One of the widely used SRIs worldwide is the normalized vegetation index NDVI $[(R_{NIR} - R_{VIS})/(R_{NIR} + R_{VIS})]$. NDVI and other SRIs have also been used to estimate chlorophyll, stay-green [51], nitrogen status, and GY in wheat and other cereals [50, 52]. For example, for continuous measurement of leaf greenness and senescence (stay-green), reflectance signature would be more informative than the index from optical chlorophyll/nitrogen leaf-clip meters. A large number of SRIs have been reported in the literature [24], some of them employing wavelengths between 1000 and 2500 nm, like the normalized water index (NWI), the normalized difference moisture index (NDMI), or the shortwave infrared water index (SWWI). These indices worked better than the NDVI to predict GY or carbon isotope discrimination ($\Delta^{13}C$) in wheat [24]. However, unlike SRIs, modeling a more significant proportion of the spectral signature can better describe the agronomic and physiological traits to be determined [26].

Thermometry is another field remote sensing technology for phenotyping [16]. Underwater scarcity conditions, g_s is reduced, reducing energy dissipation (i.e., evaporative cooling), consequently increasing leaf/canopy temperature [53]. Although thermography does not directly measure g_s , it is associated with changes in canopy temperature [54]. In this sense, thermal imaging helps characterize leaf transpiration's spatial/temporal heterogeneity and photosynthetic performance.

3 Phenomics: A Multidisciplinary Approach, from the Genome up to the Phenome

For decades, plant breeding focused its efforts on increasing yields. However, without exploring the tolerance characteristics of these genotypes, it will be challenging to estimate their behavior when the effects of climate change become more severe. Thus, understanding the biological determinants of genotypic variability in the field is one of the critical challenges for plant breeders and thus for food security [55]. Furthermore, as mentioned in several chapters of this book, the multidimensional information collected by

HTTPs and their correct interpretation will help elucidate the complex molecular mechanisms related to stress tolerance [55].

Like the gene (unit of inheritance) is part of the genome, the phenotype or single trait unit was established as “phene” [56, 57]. Later, in 1949, the word “phenome” was introduced as the sum total of extragenic, non-autoreproductive portions of the cell, whether cytoplasmic or nuclear [58]. Both words (i.e., phenome and phenotype) were defined from a genetic perspective (i.e., meaning that the phenome is what the genome influences). However, there are more complex relationships than stated in this definition. It was not until two decades ago that the term “phenomics” was introduced in analogy to genomics [59, 60].

Considering the underlying mechanisms of plant responses under adverse conditions, it would be important to consider the phenome as a complex system, with several levels of organization: from genes to the population level [61]. Since it is now possible to characterize the genome, it would be reasonable to expect it to be considered part of the phenotype [33].

Therefore, combining all available approaches (from genome to phenome) in an interdisciplinary effort (e.g., programmers, bioinformaticians, statisticians, biologists, agronomists, geneticists, physiologists) represents the next challenge for plant breeding, especially for those in charge of data matrix fusion (e.g., hyperspectral imaging and genome-wide association study (GWAS)) and trait modeling [62]. There is a unique opportunity to join efforts and expertise on common goals, such as understanding the complexity of plant responses to field-stress [63–66].

4 The Environment as a Modulator of the Phenotype and the Impact in Trait Modeling

Many genes control plant yield, but environmental conditions modulate their expression [67, 68]. For various reasons, plant breeding has focused on empirical selection for crop yield. However, this neglects traits that could be relevant for plant breeding in adverse scenarios and undermines the understanding of the $G \times E$ interaction [31, 69–71]. Furthermore, if agronomic management (M) is considered as a third relevant factor ($G \times E \times M$), the comprehension of the phenome is even more complex, especially in the field [72].

Without an adequate characterization of the environment, the generation of multidimensional information is not enough to achieve consistent models. Genotypic sensitivity to macro- and microclimate is recorded at each assessment; for example, the spectral signature of a given genotype will vary according on the environmental characteristics under which measurements are

performer. Without environmental data such as temperature (Ta), vapor pressure deficit (VPD), photosynthetic active radiation (PAR), precipitation (PP), and soil water content (SWC), model validation with data from new seasons will remain discrete [73–75]. For example, using thermal imaging data from two consecutive campaigns to estimate grain yield (GY) in 386 wheat genotypes under contrasting water regimes (i.e., full irrigation (FI) and water stress (WS)), the coefficients of determination within each condition were very dissimilar (i.e., 2011, 0.52 and 0.68, and 2012, 0.08 and 0.03, respectively) [68]. By averaging environmental (i.e., maximum daily Ta, maximum daily VPD, and SWC), phenological (i.e., days between sowing, anthesis, and mature grain), physiological (i.e., carbon isotope discrimination ($\Delta^{13}\text{C}$)), and productive variables (i.e., GY), in a cluster analysis, could establish that the two water regimes (i.e., FI and WS) did not cluster either within or between seasons (2011, 2012, 2014, 2015), nor within or between the two contrasting Mediterranean climates where said evaluated panel in Chile (Santa Rosa (sr) and Cauquenes (c)) [68] (Fig. 1). This result suggests that, in addition to characterization in terms of the amount of water applied, other factors [76] appear to play a more significant role in grouping environments with comparable characteristics.

In other words, without having to resort to combining environments to improve model statistics, likely, the coefficients of determination of a true validation (i.e., a model built with one year's measurements and validated with the next year's data) probably cannot be increased much further [76].

Spectral reflectance is closely associated with genotypic characteristics and the environmental conditions in which the individual develops and grows [16, 17, 77, 78]. Studying the impact of the environment (i.e., control, WS, heat stress, and the combination of both stresses) on the spectral signature in blueberry leaves (i.e., *Vaccinium corymbosum* L. and *V. ashei* R.), the magnitude of the differences between spectra depended on the blueberry species studied and, in the case of *V. corymbosum* cultivars, on their geographic origin (i.e., northern and southern highbush blueberry) (Fig. 2) [79].

Likewise, when comparing gas exchange and spectral signature stability as a proxy in wheat growing under the same environmental stimulus (i.e., control, WS, and the combination of WS and heat stress), the results varied according to the geographic origin where the breeder performed the selection [79]. In the case of the genotype selected under high VPD (“Fontagro” from CIMMYT, Obregón, Mexico), regardless of the environment, both g_s and spectral signature remained practically unchanged during the 3 days of evaluation (nine different leaves were assessed per day). In turn, under the same conditions, for the genotype that was selected in an ambient with periodic rainfall throughout the year (“Martha” from

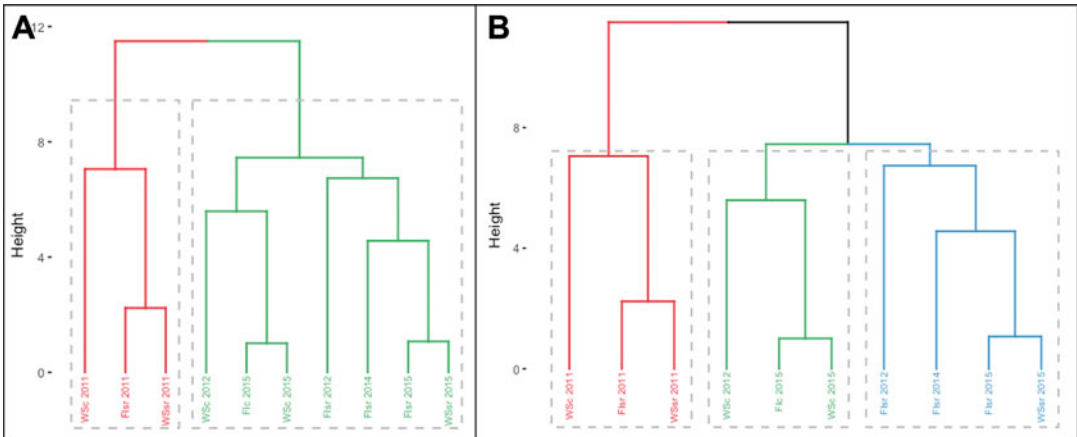


Fig. 1 Cluster dendrogram constructed with agronomic, phenological, and physiological information of 384 wheat genotypes, together with environmental information for each year and ambient studied: combination of the water regime applied (full irrigation (FI) or water stress (WS)), the experimental site (Santa Rosa (sr) or Cauquenes (c)), and growing season (2011, 2012, 2014, or 2015), expressed as a trial code in the figure. The cluster analysis was restricted to generate two (a) or three (b) groups. Data included phenological dates, productive (grain yield), physiological (carbon isotope discrimination in kernels and the stress degree day measured at the grain filling stage Z83 – SDD), and environmental information (seasonal averages of daily maximum temperature and maximum vapor pressure deficit and the soil water content between 0 and 50 cm depth). (Adapted from Romero-Bravo et al. [68])

INIA, La Estanzuela, Uruguay), g_s increased, accompanied by an evident shift of the spectral signature. Therefore, in theory, tracking changes in the spectral signature with respect to a first measurement (i.e., daily or hourly) would be a simple way to identify groups of genotypes with contrasting transpiratory behaviors (Fig. 3).

In summary, to improve statistical indicators in individual conditions (e.g., FI or WS), it will be necessary to reconsider the importance of the environment as a modulator of genotype.

5 Spectral Reflectance Data Analyses

For modeling and interpretation of predictive results, reliable information is essential [80]. However, when hundreds or thousands of genotypes are screened daily, there is a high probability of errors, especially in the field. For example, when spectral reflectance is evaluated using an optical fiber without a light source, radiometric calibration is required every 10–15 min, so the chances of inadvertently omitting this procedure are high. Something similar occurs when the fiber deviates from the canopy for a few seconds, modifying the mean spectral signature.

It is also necessary to preprocess the spectral signature and identify the attributes (wavelengths) before modeling:

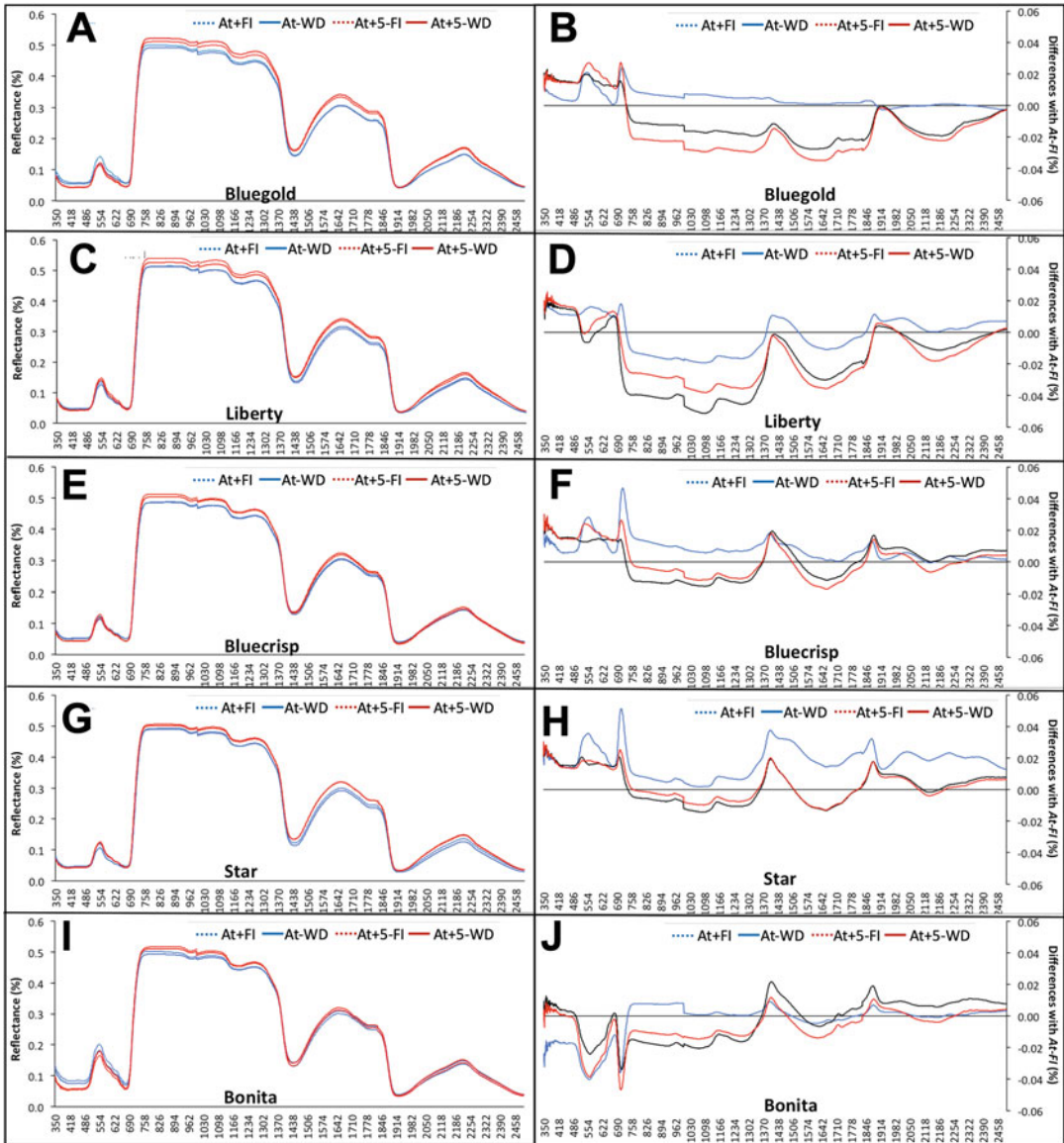


Fig. 2 Spectral signatures (a, c, e, g, i) of *V. corymbosum* (“Bluegold,” “Liberty,” “Bluecrisp,” and “Star”) and *V. ashei* (“Bonita”) growing in four ambients: *at* – *FI*, control without water deficit or heat stress; *at* – *WD*, only water deficit; *at*₊₅ – *FI*, only heat stress; and *at*₊₅ – *WD*, with water deficit and heat stress. Comparisons between control (*at* – *FI*) and each environmental condition are represented by the subtraction of reflectance at each wavelength (b, d, f, h, j) [79]

5.1 Spectral Signature Preprocessing

- (a) Data normalization or scaling: due to the mathematical functions involved, many of the machine learning algorithms are sensitive to the range of the variables; for instance, attributes whose ranges move between 0 and 100 will possess greater importance in the value of the distance to those variables

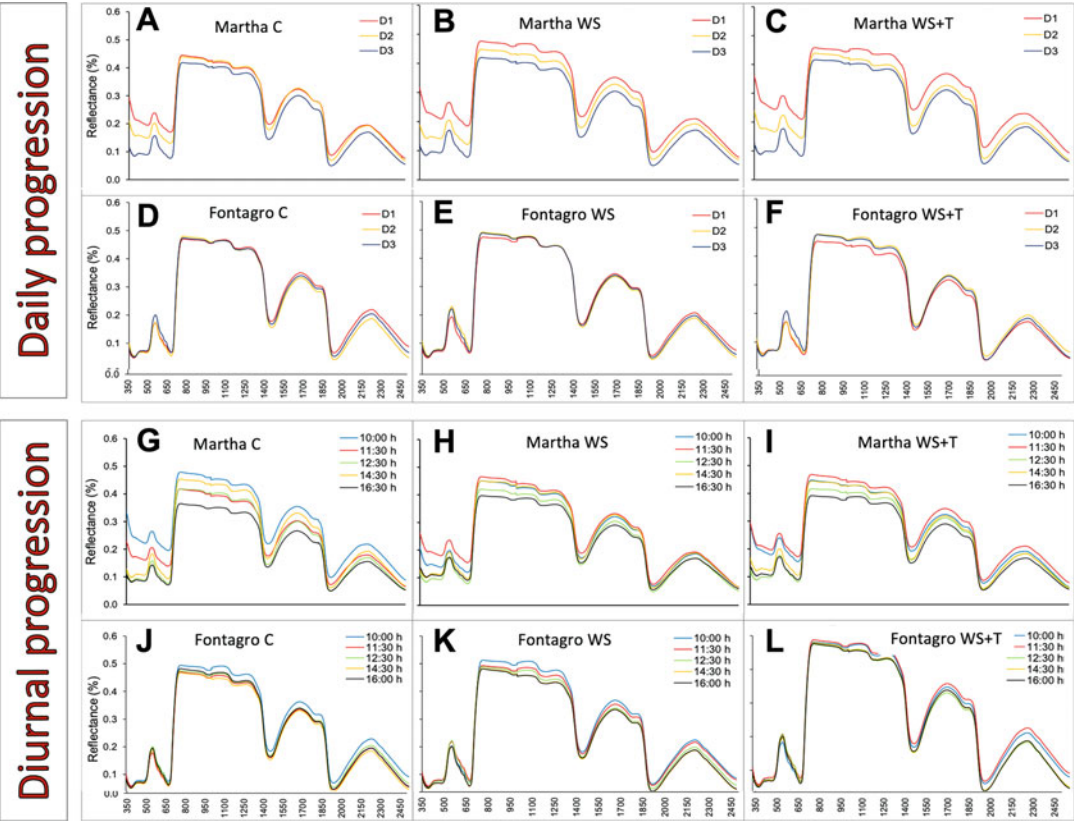


Fig. 3 Spectral signatures of leaves were recorded for three consecutive days at ~13:00 h (D1, red line; D2, yellow line; and D3, blue line) (daily progression: **a–f**) and throughout the third day of evaluation (10:00 h, blue line; 11:30 h, red line; 12:30 h, green line; 14:30 h, orange line; and 16:30 h, black line) (diurnal progression: **g–l**) in spring bread wheat genotypes “Martha” (INIA, Uruguay: **a–c, g–i**) and Fontagro (CIMMYT, Mexico: **d–f, j–l**), growing under control conditions (**c**, substrate ~75% pot water capacity and ambient temperature) (**a, d, g, j**), soil water stress (WS, substrate of ~30 of the pot water capacity and ambient temperature) (**b, e, h, k**), and the combination of water and thermal stress (WS + T, substrate ~30%, and ambient temperature increased by 5–7 °C) (**c, f, i, l**). Nine different leaves were considered for each measurement

whose ranges fall between 0 and 1. In principle, when there is no additional evidence, all the variables are assumed equal importance. Therefore, the variables often conform to a canonical form. Considering that the information collected is represented as a data matrix, in which each row represents an observation, and each column represents an attribute, several normalizations or scaling schemes can be applied [81]. Among the most relevant are as follows:

1. Normalization by area: this method normalizes the data based on the area under the curve, and all samples will have the same area [82].
2. The vector normalization, where the direction of the vector is preserved but its magnitude is set to unity.

3. Normalization by the mean, which is obtained by dividing each observation by the mean of the data matrix [83].
4. Maximum normalization, obtained by dividing each observation by the maximum absolute value of the matrix, leaving a range between -1 and 1 .
5. Range normalization, dividing each observation by the range of the matrix (maximum value $-$ minimum value).

Depending on the selected modeling criteria, it is also essential to consider that normalizations should also be applied to environmental variables.

- (b) Detection and elimination of atypical data: atypical data, also known as outliers, are observations that are too distant from the rest of the dataset. Frequently, the atypical data negatively affects the modeling. Among others, one alternative is identifying and removing the complete observation. Several families of supervised algorithms perform outlier detection, among which the local outlier factor (LOF) stands out. The LOF uses the predictor variables and the dependent variables together to determine a series of scores associated with each instance [84]. Subsequently, observations possessing LOF scores above a pre-specified threshold are identified as outliers and thus removed from the dataset.
- (c) Smoothing of the spectral signature: depending on the purposes, the smoothing of the spectral signature may reduce the spectral noise, preserving the number of predictor variables [85]. In this process, a value is transformed or replaced by the average of the values of the segment of the direct neighbors, generating a smoothed value. Here, the neighborhood concept is associated with a subset of data points whose dissimilarity (usually measured in the Euclidean distance) is slight. Different preprocessing techniques are proposed in the literature, among them (1) moving average, which generates a new value calculating the average of data within a segment of data points; (2) the Savitzky-Golay filter, which takes advantage of a moving window, selecting a subset of points adjusted based on a polynomial function; (3) the median filter, which generates a new value calculating the median of neighboring data points; and (4) the Gaussian filter, which generates a new value by calculating a weighted moving average within a subset of data points.
- (d) Derivatives of the spectral signature: the derivatives (first, second, and third order) help correct baseline effects on the spectra and solve the problem caused by overlapping bands in the spectral signature, providing a better understanding of the data [86]. Derivatives allow identifying slight variations in the

spectral sequences, tasks that can be difficult to perform through a simple inspection.

- (e) Dimensionality reduction: another critical step in data manipulation prior to the generation of the model is dimensionality reduction [87, 88]. An abridged version of the dataset is constructed by compressing the data columns, minimizing the information contained in the original dimensions. A widely used preprocessing technique is PCA, a dimensionality reduction procedure in which so-called eigenvectors are calculated to indicate the directions in which the data are dispersed.
- (f) Elimination of spectral noise: the spectral noise is inaccurate data caused by the natural presence of certain elements in the atmosphere (e.g., water, carbon dioxide), which absorb specific wavelengths [89, 90]. Because of the large number of genotypes to be screened in breeding programs, it is impossible to determine fixed or pre-established thresholds to relate the elimination of the noisy values. For example, if the measurement is taken with high relative humidity in the environment, the noise will be greater than when performed under higher VPD conditions. The criterion for eliminating noise has rarely been reported in the literature. In general, symmetrical noise cutoff is applied for all readings during the day, which may discard relevant wavelengths [80].

5.2 Attribute Selection

Modeling can consider the full spectral signature (all wavelengths) or the use of some selected wavelengths that best describe the response variable. The multi- or hyperspectral reflectance generates a large amount of data with similar reflectance information, a concept known as collinearity [91]. In other words, the neighboring spectral data contain similar information and whose high dependency provokes unpredictable performance in certain families of models [92, 93]. The detection of collinearity is usually carried out by evaluating the correlation between neighboring attributes (wavelength) (e.g., the correlation between pairs of neighbors, analysis of variance, Kappa coefficient) [92]. When the collinearity is identified, a variety of filters can be applied to reduce the adverse effects (e.g., PCA, partial least square (PLS)) and others that can select specific attributes (e.g., genetic algorithms (GA), analysis of successive steps – stepwise). Another approach is to model directly with techniques that incorporate inner methods to control the collinearity. Examples of the latter are PLS regression (PLS_r), PCA regression (PCAr), ridge regression (RR), and support vector regression (SVR). These methods allow the transformation of the original dataset and regroup the data into non-collinear vectors (PLS_r, PCAr, SVR) or limit the effect of collinearity by applying a weighting factor that reduces the contribution of collinear variables

(RR) [93]. The most frequently used procedures for attribute selection are:

- (a) Genetic algorithms (GA): the GAs is a general-purpose meta-heuristic whose goal is to find near-optimal solutions for optimization problems. The GAs are initially inspired by biological evolution, with operations based on the Darwinian principle of reproduction and survival of the individuals that best adapt to the environment. GA encodes the wavelength spectrum in the so-called chromosomes (binary vector) whose genes (binary values). In this encoding, a 0 means that the associated interval is not considered in the final solution, and a 1 means that the interval should be included. Following the biological analogy, the fittest chromosomes (the subset of wavelengths with more remarkable aptitude) are allowed to pass to the next generation, mutate, or recombine, generating offspring. After several generations (i.e., iterations of the algorithm), the information contained in the excellent candidate solutions is preserved [94]. This computational method has been widely used in the field of chemistry.
- (b) Stepwise selection: there are two types of stepwise selection, namely, forward and backward selection. On the one hand, starting from an empty solution, forward selection attempts to add attributes to a subset of characteristics incrementally. At each iteration, the method measures the quality of the solution when the analyzed attribute is included in the partial solution, identifying which of them lead to the best solution. This process is repeated until no improvements are obtained in the model or all the attributes are included in the final solution. In turn, the backward selection is analogous but this time, starting from all attributes, which are systematically removed one attribute at a time, until the model does not improve its statistical parameters. There is also the possibility of bidirectional elimination, which is a combination of the two previous methods, in which the fit of the model is tested each time a new variable is introduced or eliminated.

6 Generation of Predictive Models Through Spectral Signatures

Once the data has been preprocessed, it is necessary to select a methodology according to the requirements. For this purpose, there are three main strategies:

1. Analysis performed at the individual wavelength level.
2. Study the relationship between some wavelengths or bands of the spectrum (SRIs).

3. To model by linear and nonlinear methods, where the whole spectrum, or a part of it (attribute selection), can be exploited.

6.1 Individual Wavelengths

For the construction of new SRIs, and linear and nonlinear regression models, it would be desirable to know the degree of association of each wavelength with the variable under studied. In addition, although the association statistics could be considered discrete, what is relevant is the magnitude of the differences with the rest of the wavelengths.

6.2 SRIs

Different combinations between wavelengths have been used to generate SRIs [20, 95] to estimate phenotypic traits in plants. Commonly, SRIs are composed of wavelengths with high and low sensitivity to the expression of a given trait [96, 97]. The detection of the most relevant wavelengths for the generation of SRI can be performed by using several statistical methods, among others, including stepwise analysis, principal component analysis (PCA), or partial least squares (PLS) regression, and evaluating loading [98–100]. It has been reported that, in the face of environmental variability between seasons, which impacts the spectral signature, SRIs would be more consistent than trait modeling by linear and nonlinear regression [26, 101]. As the field spectrometer becomes cheaper, it will be possible to massively increase the use of SRI-containing bands from 1300 to 2500 nm, shown to have the potential for predicting physiological traits under challenging conditions [24]. As for SRIs databases, the SK-UTALCA software contains more than 250, which can be calculated in few steps [80].

6.3 Predictive Models

Until a decade ago, the literature focused mainly on the use of partial least squares (PLS) regression. Today, a wide range of modeling alternatives makes it challenging to compare methodologies objectively [79]. All this leads to the underutilization of methodologies or procedures that could be more effective in predicting a given trait. For example, PLS-DA stood out above all other simultaneously evaluated approaches: SRIs (255), multivariate regression methods (PCAr, PLSr, RR, and SVR), and multivariate classification algorithms (PCA linear discriminant analysis PCA-LDA, PLS discriminant analysis PLS-DA, and k nearest neighbors kNN) [26]. Although it would not be correct to compare the statistical descriptors of the above methods, the magnitude of the differences with PLS-DA indicates their potential use in plant breeding programs. In other words, it would be more efficient to identify the group of superior genotypes (elite group) instead of estimating the predicted values for each genotype.

To improve the statistical parameters of the models, it is also important to incorporate other physiological or productive traits (e.g., leaf temperature-TL, plant height, spike size, flag leaf color). Thus, when leaf temperature is incorporated into the reflectance

Table 1

Multivariate modeling (PCAr, PLSr, RR, and SVR) of grain yield as a function of reflectance (350–2500 nm) and leaf temperature (LT) in wheat plants grown under full irrigation (FI), water deficit (WS), and the combination of the databases of both environments (FI + WS)

Environment	Model	PCAr			PLSr			RR			SVR		
		RM SE	R^2	IA	RM SE	R^2	IA	RM SE	R^2	IA	RM SE	R^2	IA
FI	Reflectance	1.44	0.05	0.41	1.42	0.08	0.47	1.51	0.23	0.68	1.49	0.26	0.70
	Reflectance + LT	1.42	0.08	0.47	1.41	0.09	0.50	1.25	0.38	0.78	1.23	0.40	0.79
WS	Reflectance	0.99	0.10	0.49	0.91	0.23	0.65	1.18	0.18	0.64	1.32	0.12	0.59
	Reflectance + LT	0.98	0.11	0.51	0.89	0.26	0.66	1.07	0.18	0.66	1.18	0.17	0.64
FI + WS	Reflectance	1.49	0.17	0.56	1.47	0.19	0.59	1.31	0.45	0.82	1.28	0.46	0.82
	Reflectance + LT	1.49	0.17	0.56	1.45	0.21	0.61	1.07	0.60	0.88	1.03	0.62	0.88

Santa Rosa (Chillán, Chile), during anthesis of the 2015 season

RMSE Root mean squared error, R^2 coefficient of determination, and IA model fit index

matrix of wheat subjected to three environmental conditions (i.e., FI, WD, FI + WS), the statistics of the regression models were maintained or improved (Table 1). These results represent another example that predictive power could only improve to the extent that different databases, including those used in MAS, are merged.

Again, as shown in Table 1, predictions improved when the databases of both conditions were combined (FI + WS) [26]. This strategy, also reported by other authors [24, 102, 103], results from the larger number of observations and the larger range on the X-axis [26]. It is important to emphasize that, in general, the practice of combining environments is not consistent with the objectives of stress-oriented breeding programs, where genotypes are evaluated under specific environments (e.g., overheating, water deficit, salinity).

Although the categorical approach (PLS-DA) improved the predictive power of the traits studied, nonlinear analyses could help to improve the indicators under field conditions. An example is the subfamily of machine learning algorithms based on tree-like data structures. In computer science, tree-based learning methods (TLMs) [104] have become a valuable tool for solving complex problems [105–108], such as the degree of prediction in individual environments.

However, state of the art does not identify a single learning model as the most suitable for all situations. Therefore, there is no roadmap leading to the best learning model for most morphophysiological and physicochemical traits. In practice, each learning

method has advantages and disadvantages, and the final choice is usually a mixture of theoretical analysis and experimental results [109].

Among the TLMs showing superior potentials are logistic model trees (LMT), random forests (RF), and hierarchical neural networks (HNN). LMTs belong to a family known as decision trees, which build successive decision branches as an internal model and whose end nodes in the tree correspond to logistic functions. Instead of assigning categories at each leaf node, LMTs use logistic processes to model subsets of data. This model builds a tree for each attribute that is then combined into an ensemble method that assigns the most likely category based on a voting scheme. LMTs stop adding nodes when performance on a test set is satisfactory (the test set considers only data not used in the training process) [110].

The second type of TMLs is RFs, which belong to a more prominent family known as ensemble methods. RFs are potent models that produce successive random trees (on the order of hundreds or thousands) that are then combined to create a robust solution [111]. Each tree can produce a prediction, but combining all the trees (i.e., the forest, hence the name) will improve the predictive ability, for example, in specific environments [111, 112].

The third family of methods is hierarchical neural networks, which are mathematical models inspired by the brain's functioning. One model that performs this type of learning is the tree topology-oriented SOMs (TTOSOMs) [104]. TTOSOM succinctly represents data using structures known as neural trees trained in a series of iterations. TTOSOM assumes that neurons have plastic capabilities, allowing their internal design to be modified to represent the analyzed data. During a process known as competitive learning, the method chooses an observation from the dataset. It identifies the most similar neuron (winner neuron), which then absorbs part of the information provided by the observation. Then, the learning process is propagated through the interconnections of the tree, producing the other neurons which also absorb part of the information, as if the tree were "learning a lesson." Once the total system of neurons has been successfully trained (convergence), the method can make predictions by combining all neuron outputs into a single response. The resulting TTOSOM tree has interesting mathematical properties, such as a compact representation of the original data distribution and preservation of the underlying topology [113]. Table 2 shows an example of the predictive potential of hierarchical neural networks. In this case, TTOSOM leads to the accurate identification of elite genotypes, at least for net CO₂ assimilation, stomatal conductance, and GY.

In summary, TMLs have the advantage of being organized in simple but dynamic structures, called nodes, which communicate with other nodes through the tree's branches and, as a whole,

Table 2

Percentage of correctly classified instances for multivariate classification method (PLS-DA) and hierarchical neural network (TTOSOM), for data recollected in WS conditions

Model	Net CO ₂ assimilation	Stomatal conductance	Grain yield
PLS-DA	0.77	0.77	0.46
TTOSOM	0.90	0.88	0.92

(Santa Rosa, Chillán) during grain fill, season 2015. Santa Rosa (Chillán, Chile), during grain filling of the 2015 season

represent the global model or tree [104, 113, 114]. As the global model grows and varies according to the behavior and dispersion of the data, each new data matrix will be associated with a specific node or become a new one. With this approach, environments as we know them (e.g., IF or WS) will no longer be an arbitrary code. Thus, the predominant seasonal environmental characteristics will not only be picked up in the spectral signature of each genotype, as in the rest of the evaluated traits, but will be part of the overall model.

7 Other Considerations to Accelerate the Development of Phenomics in Plant Breeding

7.1 *Worldwide Standardization of Measurements*

It is important to concentrate efforts on determining the methodological details that are key to the coherence of the models developed, thus avoiding resorting to the combination of environments. Establishing assessment protocols that address what, how, and when is essential. For example, a practical question in annual crop improvement is whether sensors recording Ta and RH should be installed near the upper level of the canopy or use information from weather stations with sensors located 2 m above the ground.

7.2 *Phenomics in “Non-cereal” Species*

Although phenomics has evolved rapidly in cereals, work on other species groups is discrete (e.g., fruit trees, fodder crops, flowers, or forestry). In addition to the relevance for food security, cereals generally have a continuous horizontal canopy, which facilitates readings. In species such as fruit trees, vertical canopies often pose additional methodological difficulties.

7.3 *Aerial Platforms*

Aerial platforms are handy, especially in breeding programs where it is necessary to evaluate hundreds or thousands of genotypes, ideally at the same time. In the case of hyperspectral cameras (i.e., spectral signature at the pixel level), prices do not yet allow their massive implementation in field trials. Until this happens, it is suggested to develop models using reflectance recorded on the ground to be validated in aerial measurements.

7.4 Software Development

Since spectral signature modeling is still a growing field, it is essential to consider the development of customized software that allows a more efficient and faster analysis, according to the researcher's requirements. Therefore, generating a phenotyping platform (equipment + protocols + software) is essential to accelerate the generation of information that will help speed up plant breeding processes.

7.5 Joint Trials and Interdisciplinary Teams

The holistic conception of phenomics forces researchers from different fields to work on collective goals. Thus, joint trials have advantages: (1) lower operational costs; (2) statistics and interpretation of information will be based on more complex hypotheses; and (3) by involving researchers from different disciplines, potentially, it would be possible to apply for several competitive funds in the corresponding fields (e.g., ecophysiology, molecular biology, ecology, informatics), all under the same umbrella (i.e., phenomics).

7.6 Global Trials Could Compensate for the Number of Seasons Needed to Generate a Reliable Model

Model predictive ability depends, to a large extent, on the number of seasons over which a given panel is studied. Even having many years of data, predictions are likely to be more responsive in environments similar to those in which the panel has been evaluated. Thus, it is expected that, with nonlinear modeling, a panel screened globally with standardized protocols could achieve a similar or greater degree of prediction than that established in the traditional approach, but in fewer seasons.

8 Conclusions

The use of HTPPs, detailed characterization of the environment, data matrix fusion, and the nonlinear modeling approach should improve the scope of phenomics in breeding programs.

Consequently, breeding programs will be (1) more efficient (e.g., discarding unsuitable material early) [105, 115, 116]; (2) with shorter breeding cycles (fewer crosses to achieve the desired cultivar); and, therefore, (3) also increasing the probability of success at the end of the breeding process (percentage of released cultivars relative to the number of initial crosses) [15, 27].

Acknowledgments

This work was supported by the National Commission for Scientific and Technological Research CONICYT (FONDECYT 1201973 and 1180252; FONDEQUIP EQM130073 and EQM190124; FONDEF IDEA 14I10106 and 14I20106).

References

1. Ray DK, Mueller ND, West PC et al (2013) Yield trends are insufficient to double global crop production by 2050. *PLoS One* 8: e66428
2. Tilman D, Balzer C, Hill J et al (2011) Global food demand and the sustainable intensification of agriculture. *Proc Natl Acad Sci U S A* 108:20260–20264
3. Food and Agriculture Organization of the United Nations (FAO), International Fund for Agricultural Development (IFAD), World Food Programme (WFP) (2015). The state of food insecurity in the world: meeting the 2015 international hunger targets: taking stock of uneven progress. Rome
4. Dixon J, Braun HJ, Crouch J (2009) Overview: transitioning wheat research to serve the future needs of the developing world. In: Dixon J, Braun HJ, Kosina P, Crouch JH (eds) *Wheat facts and futures 2009*. CIM-MYT, México DF
5. FAOSTAT (2013) Food and Agriculture Organization of the United Nations. Statistics Division. Available online at: <http://faostat3.fao.org/>
6. Ayeneh A, Van Ginkel M, Reynolds MP et al (2002) Comparison of leaf, spike, peduncle and canopy temperature depression in wheat under heat stress. *Field Crop Res* 79:173–184
7. Azimi M, Alizadeh H, Salekdeh GH et al (2010) Association of yield and flag leaf photosynthesis among wheat recombinant inbred lines (RILs) under drought condition. *J Food Agric Environ* 8:861–863
8. Rebetzke GJ, Rattey AR, Farquhar GD et al (2012) Genomic regions for canopy temperature and their genetic association with stomatal conductance and grain yield in wheat. *Funct Plant Biol* 40:14–33
9. Challinor AJ, Watson J, Lobell DB et al (2014) A meta-analysis of crop yield under climate change and adaptation. *Nat Clim Chang* 4:287–291
10. Hernández-Barrera S, Rodríguez-Puebla C, Challinor AJ (2016) Effects of diurnal temperature range and drought on wheat yield in Spain. *Theor Appl Climatol* 129:503–519
11. Wang W, Vinocur B, Altman A (2003) Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. *Planta* 218:1–14
12. Zhang X, Cai X (2011) Climate change impacts on global agricultural land availability. *Environ Res Lett* 6:014014
13. Borrás L, Slafer GA (2008) Agronomy and plant breeding are key to combating food crisis. *Nature* 453:1177
14. Lesk C, Rowhani P, Ramankutty N (2016) Influence of extreme weather disasters on global crop production. *Nature* 529:84–87
15. Camargo AV, Lobos GA (2016) Latin America: a development pole for phenomics. *Front Plant Sci* 7:1729
16. Araus JL, Cairns JE (2014) Field high-throughput phenotyping: the new crop breeding frontier. *Trends Plant Sci* 19:52–61
17. Garbulsky MF, Peñuelas J, Gamon J et al (2011) The photochemical reflectance index (PRI) and the remote sensing of leaf, canopy and ecosystem radiation use efficiencies: a review and meta-analysis. *Remote Sens Environ* 115:281–297
18. O'Shaughnessy SA, Hebel MA, Evett SR et al (2011) Evaluation of a wireless infrared thermometer with a narrow field of view. *Comput Electron Agric* 76:59–68
19. Sinclair TR, Sinclair CJ (2010) Bread, beer and the seeds of change: agriculture's imprint on world history. CAB International, Wellingford
20. Lobos GA, Camargo A, del Pozo A et al (2017) Plant phenotyping and phenomics for plant breeding. *Front Plant Sci* 8:2181
21. Condon AG, Richards RA, Rebetzke GJ et al (2004) Breeding for high water-use efficiency. *J Exp Bot* 55:2447–2460
22. Reynolds MP, Pellegrineschi A, Skovmand B (2005) Sink-limitation to yield and biomass: a summary of some investigations in spring wheat. *Ann Appl Biol* 146:39–49
23. Araus JL, Slafer GA, Royo C et al (2008) Breeding for yield potential and stress adaptation in cereals. *CRC Crit Rev Plant Sci* 27: 377–412
24. Lobos GA, Matus I, Rodríguez A et al (2014) Wheat genotypic variability in grain yield and carbon isotope discrimination under Mediterranean conditions assessed by spectral reflectance. *J Integr Plant Biol* 56:470–479
25. Estrada F, Escobar A, Romero-Bravo S et al (2015) Fluorescence phenotyping in blueberry breeding for genotype selection under drought conditions, with or without heat stress. *Sci Hortic* 181:147–161
26. Garriga M, Romero-Bravo S, Estrada F et al (2017) Assessing wheat traits by spectral reflectance: do we really need to focus on predicted trait-values or directly identify the elite genotypes group? *Front Plant Sci* 8:280

27. Lobos GA, Hancock JF (2015) Breeding blueberries for a changing global environment: a review. *Front Plant Sci* 6:782
28. Davey JW, Hohenlohe PA, Etter PD et al (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat Rev Genet* 12:499–510
29. Jia M, Guan J, Zhai Z et al (2018) Wheat functional genomics in the era of next generation sequencing: an update. *Crop J* 6:7–14
30. Gardiner LJ, Joynson R, Hall A (2019) Chapter 3: Next generation sequencing enabled genetics in hexaploid wheat. In: Miedaner T, Korzun V (eds) *Applications of genetic and genomic research in cereals*. Woodhead Publishing, Amsterdam
31. Furbank RT, Tester M (2011) Phenomics—technologies to relieve the phenotyping bottleneck. *Trends Plant Sci* 16:635–644
32. Zhu XG, Lynch JP, Le Bauer DS et al (2016) Plants in silico: why, why now and what? An integrative platform for plant systems biology research. *Plant Cell Environ* 39:1049–1057
33. Houle D, Govindaraju DR, Omholt S (2010) Phenomics: the next challenge. *Nat Rev Genet* 11:855–866
34. Campos H, Cooper M, Habben JE et al (2004) Improving drought tolerance in maize: a view from industry. *Field Crop Res* 90:19–34
35. Richards RA (2006) Physiological traits used in the breeding of new cultivars for water-scarce environments. *Agric Water Manag* 80:197–211
36. Reynolds M, Manes Y, Izanloo A et al (2009) Phenotyping approaches for physiological breeding and gene discovery in wheat. *Ann Appl Biol* 155:309–320
37. del Pozo A, Matus I, Serret MD et al (2014) Agronomic and physiological traits associated with breeding advances of wheat under high-productive Mediterranean conditions. The case of Chile. *Environ Exp Bot* 103:180–189
38. Jedmowski C, Brüggemann W (2015) Imaging of fast chlorophyll fluorescence induction curve (OJIP) parameters, applied in a screening study with wild barley (*Hordeum spontaneum*) genotypes under heat stress. *J Photochem Photobiol B* 151:153–160
39. Mehta P, Allakhverdiev SI (2010) Characterization of photosystem II heterogeneity in response to high salt stress in wheat leaves (*Triticum aestivum* L.). *Photosynth Res* 105:49–255
40. Zivcak M, Kalaji HM, Shao HB et al (2014) Photosynthetic proton and electron transport in wheat leaves under prolonged moderate drought stress. *J Photochem Photobiol B Biol* 137:107–115
41. Ajigboye OO, Bousquet L, Murchie EH et al (2016) Chlorophyll fluorescence parameters allow the rapid detection and differentiation of plant responses in three different wheat pathosystems. *Funct Plant Biol* 43:356–369
42. Zushi K, Matsuzoe N (2017) Using of chlorophyll a fluorescence OJIP transients for sensing salt stress in the leaves and fruits of tomato. *Sci Hortic* 219:216–221
43. Pérez CEA, Rodrigues FÁ, Moreira WR et al (2013) Leaf gas exchange and chlorophyll a fluorescence in wheat plants supplied with silicon and infected with *Pyricularia oryzae*. *Phytopathology* 104:143–149
44. Jones HG (1987) Breeding for stomatal characters. In: Zeiger E, Farquhar GD, Cowan IR (eds) *Stomatal function*. Stanford University Press, Stanford
45. James RA, von Caemmerer S, Condon AT et al (2008) Genetic variation in tolerance to the osmotic stress component of salinity stress in durum wheat. *Funct Plant Biol* 35:111–123
46. Ögren E, Sjöström M (1990) Estimation of the effect of photoinhibition on the carbon gain in leaves of a willow canopy. *Plant* 181:560–567
47. Sharma DK, Andersen SB, Ottosen CO et al (2012) Phenotyping of wheat cultivars for heat tolerance using chlorophyll a fluorescence. *Funct Plant Biol* 39:936–947
48. Sharma DK, Andersen SB, Ottosen CO et al (2015) Wheat cultivars selected for high Fv/Fm under heat stress maintain high photosynthesis, total chlorophyll, stomatal conductance, transpiration and dry matter. *Physiol Plant* 153:284–298
49. Pour-Aboughadareh A, Mahmoudi M, Moghaddam M et al (2017) Agromorphological and molecular variability in *Triticum boeoticum* accessions from Zagros Mountains, Iran. *Genet Resour Crop Evol* 64:545–556
50. Aparicio N, Villegas D, Casadesus J et al (2000) Spectral vegetation indices as nondestructive tools for determining durum wheat yield. *Agron J* 92:83–91
51. Lopes MS, Reynolds MP (2010) Partitioning of assimilates to deeper roots is associated with cooler canopies and increased yield under drought in wheat. *Funct Plant Biol* 37:147–156
52. Royo C, Villegas D (2011) Field measurements of canopy spectra for biomass assessment of small-grain cereals. In: Matovic D

- (ed) Biomass-detection, production and usage. In Tech, Rijeka
53. Zia S, Spohrer K, Wenyong D et al (2011) Monitoring physiological responses to water stress in two maize varieties by infrared thermography. *Int J Agric Biol Eng* 4:7–15
 54. Jones HG (2004) Application of thermal imaging and infrared sensing in plant physiology and ecophysiology. *Adv Bot Res* 41:107–163
 55. Cobb JN, DeClerck G, Greenberg A et al (2013) Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype–phenotype relationships and its relevance to crop improvement. *Theor Appl Genet Title* 126: 867–887
 56. Johanssen W (1909) Elements of the exact theory of inheritance. Gustav Fischer, Jena
 57. Serebrovsky AS (1922) Crossing-over involving three sex-linked genes in chickens. *Am Nat* 56:571–572
 58. Davis BD (1949) The isolation of biochemically deficient mutants of bacteria by means of penicillin. *Proc Natl Acad Sci U S A* 35:1–10
 59. Dawkins R (1982) The extended phenotype: the long reach of the gene. Oxford University Press, Oxford
 60. Schork NJ (1997) Genetics of complex disease: approaches, problems, and solutions. *Am J Respir Crit Care Med* 156:S103–S109
 61. Marshall-Colon A, Long SP, Allen DK et al (2017) Crops in silico: generating virtual crops using an integrative and multi-scale modeling platform. *Front Plant Sci* 8:786
 62. Feng H, Guo Z, Yang W et al (2017) An integrated hyperspectral imaging and genome-wide association analysis platform provides spectral and genetic insights into the natural variation in rice. *Sci Rep* 7:4401
 63. Thrall PH, Oakeshott JG, Fitt G et al (2011) Evolution in agriculture: the application of evolutionary approaches to the management of biotic interactions in agro-ecosystems. *Evol Appl* 4:200–215
 64. Reynolds M, Langridge P (2016) Physiological breeding. *Curr Opin Plant Biol* 31:162–171
 65. Dhondt S, Wuyts N, Inzé D (2013) Cell to whole-plant phenotyping: the best is yet to come. *Trends Plant Sci* 18:428–439
 66. Ghanem ME, Marrou H, Sinclair TR (2015) Physiological phenotyping of plants for crop improvement. *Trends Plant Sci* 20:139–144
 67. Sadras VO, Reynolds MP, De la Vega AJ et al (2009) Phenotypic plasticity of yield and phenology in wheat, sunflower and grapevine. *Field Crop Res* 110:242–250
 68. Romero-Bravo S, Méndez-Espinoza AM, Garriga M et al (2019) Thermal imaging reliability for estimating grain yield and carbon isotope discrimination in wheat genotypes: importance of the environmental conditions. *Sensors* 19:2676
 69. Yang N, Lu Y, Yang X et al (2014) Genome wide association studies using a new nonparametric model reveal the genetic architecture of 17 agronomic traits in an enlarged maize association panel. *PLoS Genet* 10:e1004573
 70. Großkinsky DK, Svensgaard J, Christensen S et al (2015) Plant phenomics and the need for physiological phenotyping across scales to narrow the genotype-to-phenotype knowledge gap. *J Exp Bot* 66:5429–5440
 71. Rahaman MM, Chen D, Gillani Z et al (2015) Advanced phenotyping and phenotype data analysis for the study of plant growth and development. *Front Plant Sci* 6:619
 72. Porter JR, Christensen S (2013) Deconstructing crop processes and models via identities. *Plant Cell Environ* 36:1919–1925
 73. Leinonen I, Grant OM, Tagliavia CPP et al (2006) Estimating stomatal conductance with thermal imagery. *Plant Cell Environ* 29: 1508–1518
 74. Grant OM, Tronina L, Jones HG et al (2007) Exploring thermal imaging variables for the detection of stress responses in grapevine under different irrigation regimes. *J Exp Bot* 58:815–825
 75. Prashar A, Jones H (2014) Infrared thermography as a high-throughput tool for field phenotyping. *Agronomy* 4:397–341
 76. Reynolds MP, Quilligan E, Aggarwal PK et al (2016) An integrated approach to maintaining cereal productivity under climate change. *Glob Food Sec* 8:9–18
 77. Garriga M, Retamales JB, Romero-Bravo S et al (2014) Chlorophyll, anthocyanin, and gas exchange changes assessed by spectroradiometry in *Fragaria chiloensis* under salt stress. *J Integr Plant Biol* 56:505–515
 78. Hernández J, Lobos GA, Matus I et al (2015) Using ridge regression models to estimate grain yield from field spectral data in bread wheat (*Triticum aestivum* L.) grown under three water regimes. *Remote Sens* 7:2109–2126
 79. Lobos GA, Escobar-Opazo A, Estrada F et al (2019) Spectral reflectance modeling by wavelength selection: studying the scope for blueberry physiological breeding under

- contrasting water supply and heat conditions. *Remote Sens* 11:329
80. Lobos GA, Poblete-Echeverría C (2017) Spectral Knowledge (SK-UTALCA): software for exploratory analysis of high-resolution spectral reflectance data on plant breeding. *Front Plant Sci* 7:1996
 81. Rinnan Å, Van Den Berg F, Engelsen SB (2009) Review of the most common pre-processing techniques for near-infrared spectra. *TrAC-Trends Anal Chem* 28:1201–1222
 82. Zheng X, Ren H, Qin Q et al (2015) Retrieval of canopy water content using a new spectral area index method. Paper presented in 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, 24 July 2015
 83. Muhammed HH (2005) Hyperspectral crop reflectance data for characterising and estimating fungal disease severity in wheat. *Biosyst Eng* 9:19–20
 84. Breunig MM, Kriegel HP, Ng RT et al (2000) LOF: identifying density-based local outliers. *ACM Sigmod Rec* 29:93
 85. Rollin EM, Milton EJ (1998) Processing of high spectral resolution reflectance data for the retrieval of canopy water content information. *Remote Sens Environ* 65:86–92
 86. Tsai F, Philpot W (1998) Derivative analysis of hyperspectral data. *Remote Sens Environ* 66:41–51
 87. Harsanyi JC, Chang CI (1994) Hyperspectral image classification and dimensionality reduction: an orthogonal subspace projection approach. *IEEE Trans Geosci Remote Sens* 32:779–785
 88. Roweis ST, Saul LK (2000) Nonlinear dimensionality reduction by locally linear embedding. *Science* 290:2323–2326
 89. Curtiss B, Goetz AFH (2001) Field spectrometry: techniques and instrumentation. Analytical Spectral Devices Inc. <http://www.asdi.com/Field%20Spectroscopy-screen.pdf>. Accessed 20 Jan 2019
 90. Clevers JG, Kooistra L, Schaepman ME (2008) Using spectral information from the NIR water absorption features for the retrieval of canopy water content. In *J Appl Earth Obs Geoinf* 10:388–397
 91. Ravikanth L, Jayas DS, White ND et al (2017) Extraction of spectral information from hyperspectral data and application of hyperspectral imaging for food and agricultural products. *Food Bioproc Tech* 10:1–33
 92. Dormann CF, Elith J, Bacher S et al (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36:27–46
 93. Hastie T, Tibshirani R, Friedman J et al (2005) The elements of statistical learning: data mining, inference and prediction. *Math Intell* 27:83–85
 94. Davis L (1991) Handbook of genetic algorithms. Van Nostrand Reinhold, New York
 95. Inoue Y, Peñuelas J, Miyata A et al (2008) Normalized difference spectral indices for estimating photosynthetic efficiency and capacity at a canopy scale derived from hyperspectral and CO₂ flux measurements in rice. *Remote Sens Environ* 112:156–172
 96. Bannari A, Morin D, Bonn F et al (1995) A review of vegetation indices. *Remote Sens Environ* 13:95–120
 97. Gitelson AA, Merzlyak MN (1996) Signature analysis of leaf reflectance spectra: algorithm development for remote sensing of chlorophyll. *J Plant Physiol* 148:494–500
 98. Thenkabail PS (2001) Optimal hyperspectral narrowbands for discriminating agricultural crops. *Remote Sens Rev* 20:257–291
 99. Mahlein AK, Rumpf T, Welke P et al (2013) Development of spectral indices for detecting and identifying plant diseases. *Remote Sens Environ* 128:21–30
 100. Elsayed S, Elhoweity M, Schmidhalter U (2015) Normalized difference spectral indices and partial least squares regression to assess the yield and yield components of peanut. *Aust J Crop Sci* 9:976–986
 101. Jackson RD, Slater PN, Pinter PJ (1983) Discrimination of growth and water stress in wheat by various vegetation indices through clear and turbid atmospheres. *Remote Sens Environ* 13:187–208
 102. Aparicio N, Villegas D, Araus JL et al (2002) Relationship between growth traits and spectral vegetation indices in durum wheat. *Crop Sci* 42:1547–1555
 103. Royo C, Aparicio N, Villegas D et al (2003) Usefulness of spectral reflectance indices as durum wheat yield predictors under contrasting Mediterranean conditions. *Int J Remote Sens* 24:4403–4419
 104. Astudillo CA, Oommen BJ (2014) Topology-oriented self-organizing maps: a survey. *Pattern Anal Applic* 17:223–248
 105. Cerpa N, Bardeen M, Astudillo CA et al (2016) Evaluating different families of prediction methods for estimating software project outcomes. *J Syst Softw* 112:48–64
 106. Astudillo CA, Poblete J, Resta M et al (2016) A cluster analysis of stock market data using hierarchical SOMs. In: Kang BH, Bai Q (eds)

- Advances in artificial intelligence: 29th Australasian joint conference. Springer, Hobart
107. Maldonado G, Astudillo CA, Riadi G et al (2017) Predicting the stability of human lysozyme mutants using the tree-based classifier TTOSOM. *Chemom Intell Lab* 162:65–72
 108. Nuñez H, Maldonado G, Astudillo CA (2018) Semi-supervised regression based on tree SOMs for predicting students' performance. In: 9th International conference on pattern recognition systems (ICPRS 2018), Valparaiso, 22 May 2018
 109. Alpaydin E (2004) Introduction to machine learning. MIT Press, Cambridge, MA
 110. Landwehr N, Hall M, Frank E (2005) Logistic model trees. *Mach Learn* 59:161–205
 111. Breiman L (2001) Random forests. *Mach Learn* 45:5–32
 112. Biau G (2012) Analysis of a random forests model. *J Mach Learn Res* 13:1063–1095
 113. Astudillo CA, Oommen BJ (2011) Imposing tree-based topologies onto self-organizing maps. *Inf Sci* 181:3798–3815
 114. Astudillo CA, Oommen BJ (2015) Pattern recognition using the TTOCONROT. In: Ali M, Kwon Y, Lee CH, Kim J, Kim Y (eds) International conference on industrial, engineering and other applications of applied intelligent systems, vol 9101. Springer, Cham
 115. Tardieu F, Tuberosa R (2010) Dissection and modelling of abiotic stress tolerance in plants. *Curr Opin Plant Biol* 13:206–212
 116. Brennan JP, Martin PJ (2007) Returns to investment in new breeding technologies. *Euphytica* 157:337–349



Chapter 14

Designing Experiments for Physiological Phenomics

Addie Thompson, Michael Kantar, and Katy Rainey

Abstract

Phenomics has emerged as the technology of choice for understanding quantitative genetic variation in plant physiology and plant breeding. Phenomics has allowed for unmatched precision in exploring plant life cycles and physiological patterns. As new technologies are developed, it is still vital to follow best practices for designing and planning to be able to fully exploit any experimental results. Here we describe the basic – but sometimes overlooked – considerations of a phenomics experiment to help you maximize the value from the data collected: choosing population and location, accounting for sources of variation, establishing a timeline, and leveraging ground-truth measurements.

Key words Remote phenotyping, Automated data collection, Experimental workflow, Correlated measures, Phenomics, Experimental design

1 Introduction

1.1 Defining a Phenotype and Why We Measure

A goal of natural science has always been to understand the way the world works, and in order to understand the world, things need to be measured. These measurements are essential to develop an expectation for the natural environment and human-mediated environment. Historically, humanity has desired different things of the environment; one of the major goals has been to produce food. Plants have been central to food production since the advent of agriculture [1]. A plant is an organism that is used to understand a specific species or a specific genotype in order to gain insight into how genotype, environment, and management interact. Plants are central to agriculture and the natural environment. Measurements are used to develop expectations which are foundational to improving populations based on the ever-shifting landscape of human goals.

To reach human goals, science uses specialized terms, with the same words often meaning different things even between sub-disciplines [2]. Clarifying definitions and the goals of terminology is important so that objectives can more readily be achieved.

One of the most common terms is phenotype, which can be defined as a measurable characteristic that is the physical outcome of the combination of the genotype and the environment that an organism lives in [3]. New high-throughput phenotypes allow for an unprecedented way to make measurements; these techniques greatly increase the amount of data [4]. This new data can be data that wouldn't have been previously collected, or it can increase the precision of the data that is collected [5].

However, the phenotypes that are commonly measured can be the outcome of physiological important genetic pathways or simply correlated with important genetic pathways. Frequently, this means that measurements are conducted based on what is possible rather than what is most relevant to the question; this is most frequently a correlated phenotype [6].

1.2 History of Image-Based Data Collection

Image analysis and remote sensing have been a technique used in science since the mid-nineteenth century [7]. The first large-scale landscape photos were taken in the 1850s from balloons either suspended or free-floating. These simple black and white photographs formed the basis for the development of analytical techniques that have been used in a diverse array of fields from sociology to oceanography to agriculture. The basic workflow for processing images has to do with spectral differentiation, radiometric differentiation, spatial differentiation, and geometric transformations.

The first use of aerial vehicles for image collection was World War I. The success of pictures is based on optical elements of the camera as well as the conditions on the day pictures are taken. Historically, different types of film, different shutter speeds, and different types of optics led to different visual contrasts and thus differential ability to identify elements on the ground. The ability to ensure accuracy is dependent on being able to identify geometry associated with each image and correct for individual geometric differences between images.

The advent of orbital imagery allowed for a different spatial scale than had previously been available. The first satellite systems designed to monitor land resources (e.g., crops) were launched in the early 1970s. New satellites have been sent up with regularity since that time, each new satellite has increased the range in the electromagnetic spectrum that has higher resolution to explore different features across earth. The temporal resolution of satellites is limited by the rate at which they orbit the earth.

The use of unmanned aerial vehicles allowed for a new wide range of sites to be available for study and for much higher resolution both spatially and temporally for each site [8]. An additional way to achieve temporal and spatial resolution is to mount cameras on tractors and measure plants within rows [9]. These have been expanded to include a wide range of sensors to measure a large number of traits [10]. It is possible to have fixed cameras mounted

to view entire fields that can take time series images at multiple resolutions to understand spatial and temporal variation.

1.3 Approaches from Earth Science, Geography, and Ecology

Recently, high-throughput phenotyping has focused on image data. Once images have been created, it is necessary to be able to understand and interpret them, and this can be challenging [11]. Image analysis has a long-standing history of use in many fields, providing a large opportunity for interdisciplinary learning. Three disciplines that have long used observational data to explore image analysis to understand biophysical processes are geography/GIS, earth science, and ecology.

Remote sensing has the goal of gaining information about land surfaces and bodies of water based on long-distance measurements. As the number of measurements increased and more types of features were interrogated, algorithms were developed in order to classify these features. The different types of imagery are varied and lead to the classification of different types of features. These techniques are based on contrast between features, on different visual spectra. Pictorial images are generally transitioned to digital formats which subdivide the images into grids of equal sizes. Classifying land cover has been a major source of method development in remote sensing and has been very useful for understanding large-scale landscape changes over time and space. With respect to plant material, the initial challenges were with respect to vegetation or no vegetation; then to what type of vegetation (forest, grassland), life history of plant material, and specific type of plant material; and then finally to ask specific questions about the plant material. This final approach is of most use to plant breeding and agronomy, as it allows for the prediction of yield, disease incidence, insect incidence, growth rate, and abiotic tolerance.

2 Materials

2.1 Options for Plants

1. One or a few genotypes across multiple other non-genetic treatment levels.
2. A small number of contrasting genotypes, species, or mutant vs. wild-type.
3. A modest number (20–50) of diverse genotypes for surveying variation, estimating heritability, and calibrating models to predict phenotypes from correlated traits.
4. A moderate number (100–300) of genotypes for mapping quantitative trait loci (QTL) in a segregating biparental mapping population or recombinant inbred line population.

5. A larger number (300–1000+) of diverse genotypes or breeding lines for conducting genome-wide association studies (GWAS) or genomic prediction.

2.2 Potential Facilities

1. Growth chamber with carefully controlled temperature, lighting, and humidity.
2. Phenotyping greenhouse with well-controlled temperature, lighting, and humidity; plants may be on conveyor belts to minimize spatial variation.
3. Standard greenhouse, level of control of temperature, lighting, and humidity can vary; plants in pots on the floor or on shelves or sometimes directly in the ground.
4. Hoop house or hot house, little to no climate control, clear plastic-covered shelter to increase heat and extend growing season.
5. Research field plots on your own land (university farm, industry field test site).
6. On-farm field trial using land maintained by a farmer.

2.3 Example Data Collection Platforms

1. Fixed-point installations: one or many networked cameras mounted in place in a field, greenhouse, or chamber.
2. Balloon: tethered balloon with sensors fixed to the bottom.
3. Satellite: sensors carried in orbit in space and used to image the earth.
4. Tower: tall pole-like structure in a field, greenhouse, or chamber with sensors at the top looking out/down.
5. Imaging “box” or “imaging tower”: a self-contained structure, usually in a greenhouse or lab, that plants enter (either automatically on a conveyor belt or manually) with controlled lighting and fixed-position sensors.
6. Gantry: moving suite of sensors on rails or cables that moves over a fixed area.
7. Tractor, lawnmower, or all-terrain vehicle: can drive sensors around a field, but must have special row spacing to drive in the field or are limited by plant height.
8. Cart: low-cost usually human-powered set of sensors on wheels to push over field plots.
9. Modified high sprayer: enables driving into a field with normal row spacing and taller plant height.
10. Field robot: drives in between rows of plants carrying sensors at potentially varying angles and heights to collect in-canopy data.

11. Unmanned aerial systems (UAS): fixed-wing or roto-copter drone flying over a field site to collect above-canopy data on crop performance or environmental conditions.

2.4 Potential Sensors

1. Black and white: very basic image that captures light and dark contrasts; can be used for size/shape/height (collectively, geometry).
2. RGB, or red-green-blue: most common type of camera imagery that captures what the human eye sees; can be used for color if well-calibrated, as well as geometry; overlapping images allow for approximation of depth and height.
3. Multispectral: a few (usually 2–7) spectral bands that are often used to calculate vegetative indices, indicative of traits like chlorophyll content.
4. Hyperspectral: many (often 100 or more) high-resolution spectral bands that are used to explore and model relationships between spectral reflectance and plant health, status, and biochemical composition.
5. Thermal: infrared measurements to estimate temperature.
6. LiDAR: sensor device containing spinning lasers that bounce off structures and measure time-to-return as a way to map complex shape and structure as point clouds in three-dimensional space.

2.5 Ground Reference Tools

1. Basic hand tools, such as a height stick, calipers, and scale (*see Note 1*).
2. Chlorophyll meter, SPAD, or other handheld tool for plant greenness.
3. Gas exchange system to measure photosynthesis and respiration.

3 Methods

3.1 Consider the Starting Materials

Careful consideration of the materials to be phenotyped is critical to success with physiological phenomics. The objectives of the experiment dictate the genotypes of interest, for example, if a contrast is to be made between a mutant and wild-type plant. Often, multiple genotypes (lines or varieties) will be compared to investigate a trait of interest. The choice of genotypes, populations, or panels will depend on your specific objectives. Some questions to consider:

1. What variation exists for your trait(s)? If you are the very first to measure a particular phenotype, it is usually best to start with replications of a few extremely diverse lines to estimate variation within and among genotypes and calculate heritability.

2. What is the genetic architecture of the trait(s) of interest? If you know your trait is highly quantitative and likely to be controlled by many genes with small effects, you will need to use a greater number of lines and likely more replications in order to accurately estimate the effects of any one loci in a mapping experiment, for example.
3. What is the heritability? Traits with low heritability and/or those that are dramatically affected by the environment or soil conditions will also challenge your ability to obtain accurate estimates. Highly heritable traits can be measured using fewer replications and/or locations. The relative importance of high heritability will vary depending on your experimental goals, but generally it is advisable to maximize the ratio of the genetic variation to the total variation in an experiment. This increases the likelihood that the experimental results can be replicated and decreases the apparent level of “noise” in measuring the trait. If phenotyping is being conducted as part of a breeding program, sufficiently high heritability becomes essential to ensure efficient gain from selection. If heritability is low for your phenotype when measured directly, are there correlated traits (related to or predictive of your phenotype of interest) that are more highly heritable? If so, this would be beneficial to consider.
4. What effect size are you looking for? If you hope to capture very small effect sizes, you will need to increase your replications (for experiments involving direct comparisons) or number of genotypes (for genetic mapping projects). Increasing the number of genotypes will help to offset the Beavis effect [12, 13] of over-estimating small effects and under-estimating the number of loci involved.
5. In which genotypes does variation exist for your trait(s) of interest? The starting population will determine what contrasts you are able to capture. For example, selecting a population of highly elite lines will make it very difficult to identify significant genetic contributions to yield, as the genetic variation and phenotypic variance among the starting set will be low. In fact, physiological phenomics often requires collecting yield data on low-yielding material, which is not conventional to agronomic research methods. Diversity panels often consist of landraces which have been selected for yield and may be reduced for genetic and phenotypic variance for yield-associated traits as well. In a biparental population, it is possible to have transgressive segregation (i.e., phenotypic values in the offspring outside the range of the parents), but you will still be limited to identify only the loci that are segregating.

3.2 Consider the Location and Environment

How and where you intend to grow your plants will impact your experimental planning and design. Your choice of environment and/or platform will impact – and be impacted by – your chosen genotypes, population, or panel.

1. Will you be growing your material in a growth chamber, greenhouse, or field space? Often you will sacrifice some level of experimental control in exchange for “real-world” relevance at scale.
2. How many lines will realistically fit in your space, given the number of replications you may need to estimate the trait? You will have to consider the tradeoff between increasing replication and increasing the number of genotypes observed, with the constraint of space in mind.
3. What costs are involved? You may have costs associated with facility rental (chamber, greenhouse, field), equipment use (planter/tractor/harvester, phenotyping equipment), supplies (pots, soil, fertility, pesticides, stakes), technical support, and additional labor for maintenance and data collection. Given these costs, how many pots or plots can you afford?
4. How many pots or plots can you properly manage (planting, watering, fertilization, weeding, etc.)? A smaller well-managed experiment is often better than a larger poorly managed one.
5. At what stage(s) of growth and in what tissue(s) do you intend to phenotype your plants? If you need to grow your material to reproductive or maturity stages, are your chosen lines able to mature in your environment? There are many environmental and logistical considerations for this question. For example, some genotypes and/or species require specific day lengths in order to flower, some require vernalization (a cold period), and some require the accumulation of a certain amount of heat units prior to frost. Mature plant size may be an additional concern for indoor facilities.

3.3 Consideration of Non-treatment Sources of Variation in Multi-environment Yield Trials

Decide whether and how to account for them or eliminate them (Fig. 1).

1. Make valid assumptions about what variation is important to your question. For example, if you are exploring a specific abiotic stress (e.g., drought), make sure you can capture the relevant variation associated with the expression of that stress while eliminating or accounting for variation that may confound your conclusions, such as changes in soil type or elevation throughout a field or air circulation patterns in the greenhouse or chamber.

Sources of Non-treatment Variance in Agronomic Research

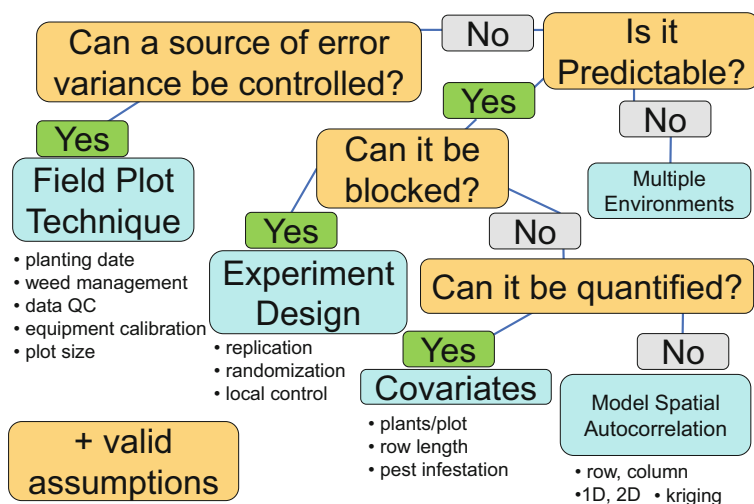


Fig. 1 Minimizing error in a large-scale phenotyping experiment by accounting for non-treatment sources of variation

2. Can the source of error variation be controlled? If so, this is incorporated into your technique including decisions about management such as preparation of the seed bed, seeding rate, plot size (often dictated by whether yield is being considered), and planting date. You can eliminate error with properly calibrated equipment and data quality control protocols. Check phenotypes as soon as possible after collection so that extreme values (outliers) can be re-evaluated to determine whether they are errors or true rare variants. If your experiment requires destructive ground reference samples, then plots can have multiple rows to separate yield rows from sampling rows while retaining borders.
3. Is the source of error variation predictable? If you can predict a source of error but you can't control it, then you can devise a way to account for it. If a source of variance is unpredictable, such as rainfall, then you should conduct your experiments in multiple environments, both to eliminate risk and to obtain a valid estimate of your treatment effect/model your environmental variance.
4. Can the source of error variation be blocked? Blocking for effects such as soil gradients and slopes and treatment interactions are your elements of experimental design. The three basic principles of experimental design are replication, randomization, and local control.

5. Can the source of error variation be quantified? Predictable sources of variation that can't be controlled or blocked can be quantified and used as covariates in statistically adjusted phenotypes (*see Note 2*). These include unintended phenomena such as pest infestation and variation in seedling emergence and plot size. High-throughput phenotyping platforms provide new methods of accounting for these sources of error, such as collecting stand counts, measuring exact row lengths, and detecting stress from changes in growth rate and color.

3.4 Plan Out the Experiment in Advance with a Timeline

After considering the genotypes and environment, you will need to consider the timing of your measurements. A detailed timeline will keep all individuals involved in data collection informed of expectations and necessary planning.

1. When will measurements be taken? Determine what phenotypes are most important and at what stages. Draw out a calendar that contains all phenotyping activities, what labor and equipment are involved in each step, and what priority different activities have in case there is a conflict due to weather conditions or other unforeseen events.
2. Decide whether you will time your measurements by days after planting, growth stage, or thermal time. Days after planting may be convenient for a large experiment, but timing by growth stage might make your measurements of physiological phenomena more comparable (and decrease the impact of flowering time on your trait). Using thermal time (e.g., growing degree days or GDD) will be similar to days after planting except that it accounts for the fact that warm, sunny days will lead to more growth than cool, cloudy days. Thermal time allows you to space measurements uniformly by plant growth rather than calendar days to avoid wasting resources (measuring too many times during slow growth) or missing key growth stages (not measuring enough during rapid growth).
3. What is the temporal resolution that you need? Some traits are only visible or relevant during a particular time frame – for example, flowering. Some traits like stand count may only need to be taken once during the season. Others, like biomass or height, accumulate at varying rates over the course of the growth cycle. Accumulation traits often follow a sigmoidal-type curve that can be used to model changes throughout the season. Knowing this, it is often more important to capture multiple timepoints during the rapid growth cycle mid-season than either the early or late stages. Time of day may also be crucial to your trait(s). For example, spectral data is often best acquired within 1–2 h of solar noon at your location. This may impact how many plants you are able to measure during a constrained time frame.

4. What is the spatial resolution that you need? For image-based phenotyping, higher resolution enables more accurate estimation of more traits. However, high-resolution sensors and platforms are often cost-prohibitive and create tremendous quantities of data that may not be necessary for the questions at hand. On the other hand, some specialized image techniques may rely on a minimum number of pixels in a defined region in order to correctly match or estimate patterns.
5. What other data need to be collected at the same time to add value? If you are taking image-based measurements, what manual measurements (“ground truth”) might you need to take for modeling? Are there samples you should collect at the same time? Consider taking note of growth stage, recording weather and management data, and collecting soil information.

3.5 Phenomic Inference: A Multi-objective Framework for Evaluating Unmanned Aerial Systems (UAS) Phenotyping Capabilities

Aim Combine quantitative genetics and physiological growth analysis to make inferences for new remote sensing capabilities for yield prediction.

Given A panel or population of lines (*see Note 3*) exhibiting:

- *Temporal quantitative variation in physiological and/or longitudinal phenotypes (i.e., biomass), replicated and assessed from the ground and remotely using UAS*
- *A calibration panel of a randomly chosen subset of lines*
- *Quantitative genetic and phenotypic variation for grain yield potential (see Note 4) measured empirically from all plots*
- *Control of factors that confound interpretation, such as flowering time, height, or population structure*
- *Measures of development and phenology, i.e., growth stages*
- *Relevant physiological measurements to interpret results, i.e., gas exchange parameters (see Note 5)*
- *High-density genome-wide markers to estimate kinship among lines (K matrix)*

Calculate

- Summary phenotypic parameters describing development via multivariate mixed models, i.e., BLUPs and BLUEs (*see Note 6*)
- Genetic and environmental variances for summary parameters via variance decomposition using the K matrix

Determine

- Genome-wide analyses, including GWAS outputs (QTL, genetic architecture, and associations of genomic regions throughout development), and assessment of the contribution of the

UAS-derived secondary traits to predictive ability for genomic prediction [14]

- Quantitative genetic properties for application to breeding including genetic correlations among traits, other measures of association to yield, and narrow-sense heritability
- Remote sensing prediction equations via training/validation approaches (*see* **Note 7**)

4 Notes

1. This is a small subset of potential materials for ground-truth data collection, depending on experimental goals. For example, prediction of chlorophyll, N status, and other molecules will require reference samples and wet chemistry.
2. New advances in space-time analysis are making spatiotemporal methods accessible. There are two major methods of approach: descriptive and dynamic [15]. Descriptive modeling seeks to characterize the mean and covariance function; this type of modeling builds on spatial statistics (e.g., kriging [16], reviewed in [17]) and tends to be very useful when processes are not well understood. Dynamic models take a conditional approach that attempts to model the mechanistic real-world process that allows for more causal inference.
3. A recommended approach to creating a phenomic inference panel is to subset a MAGIC or NAM population by using published information to control for factors that confound interpretation, for example, restricting to lines that flower within a narrow time frame. The restricted subset can be randomly sampled to further reduce the panel size. If possible, confirm that the subset panel retains a high variance for grain yield potential. A minimum of approximately 400 lines replicated twice per environment is preferable. MAGIC and NAM populations are ideal because they are unselected mapping population that are segregating for grain yield potential and a range of associated traits.
4. Or another highly quantitative and expensive trait of interest such as resistance to biotic or tolerance to abiotic stress, etc.
5. If sampling or phenotyping must be spread over several days, consider phenotyping certain lines repeatedly as a form of control.
6. Summary parameters are ideally derived from a growth analyses (i.e., logistic growth curve) and are adjusted by thermal time or other measures of development and might include summaries over growth stages and calculations of growth rate, etc.

7. Training and validation of remote sensing prediction equations for phenomic traits is necessary when ground reference sampling is too expensive to collect on all plots, but a large number of treatments are needed for genetic analyses. In addition to your replicate blocks with plots of all treatments/lines, plant a calibration block of a randomly selected subset of lines from your phenomic inference panel. Collect ground reference data (i.e., biomass) from the calibration block on every sampling date. Collect ground reference data from a randomly chosen subset of plots in the full-panel replicate blocks on some or all sampling dates. Develop your remote sensing prediction model using ground reference data from the calibration block, and then validate that the model accurately predicts ground reference values in the full-panel replicate blocks. Then, apply your best model to quantify your phenomic trait phenotypes for all plots from UAS data. Ground reference phenotypes and remotely determined phenotypes can be used in separate GWAS analyses as a means of validation.

References

1. Harlan JR (1971) Agricultural origins: centers and non-centers. *Science* 174(4008):468–474
2. Myers G (1991) Lexical cohesion and specialized knowledge in science and popular science texts. *Discourse Process* 14(1):1–26
3. Chitwood DH, Topp CN (2015) Revealing plant cryptotypes: defining meaningful phenotypes among infinite traits. *Curr Opin Plant Biol* 24:54–60
4. Araus JL, Cairns JE (2014) Field high-throughput phenotyping: the new crop breeding frontier. *Trends Plant Sci* 19(1):52–61
5. Haghighattalab A, Pérez LG, Mondal S et al (2016) Application of unmanned aerial systems for high throughput phenotyping of large wheat breeding nurseries. *Plant Methods* 12(1):35
6. Cabrera-Bosquet L, Crossa J, von Zitzewitz J et al (2012) High-throughput phenotyping and genomic selection: the frontiers of crop breeding converge. *J Integr Plant Biol* 54(5):312–320
7. Campbell JB, Wynne RH (2011) *Introduction to remote sensing*. Guilford Press, New York
8. Bagheri N (2017) Development of a high-resolution aerial remote-sensing system for precision agriculture. *Int J Remote Sens* 38(8–10):2053–2065
9. Ruixiu S, Wilkerson JB, Wilhelm LR et al (1989) A microcomputer-based morphometer for bush-type plants. *Comput Electron Agric* 4:43–58
10. Andrade-Sanchez P, Gore MA, Heun JT et al (2014) Development and evaluation of a field-based high-throughput phenotyping platform. *Funct Plant Biol* 41(1):68–79
11. Kelly D, Vatsa A, Mayham W et al (2016) An opinion on imaging challenges in phenotyping field crops. *Mach Vis Appl* 27(5):681–694
12. Beavis WD, Grant D, Albertsen M et al (1991) Quantitative trait loci for plant height in four maize populations and their associations with qualitative genetic loci. *Theor Appl Genet* 83:141–145
13. Beavis WD (1998) QTL analyses: power, precision, and accuracy. *Mol Dissect Complex Traits* 1998:145–162
14. van Eeuwijk FA, Bustos-Korts D, Millet EJ et al (2019) Modelling strategies for assessing and increasing the effectiveness of new phenotyping techniques in plant breeding. *Plant Sci* 28:23–39
15. Wikle CK, Zammit-Mangion A, Cressie N (2019) *Spatio-temporal statistics with R*. CRC Press, Milton
16. Krige DG (1951) A statistical approach to some basic mine valuation problems on the Witwatersrand. *J South Afr Inst Min Metall* 52(6):119–139
17. Cressie N (1990) The origins of kriging. *Math Geol* 22(3):239–252



Design Considerations for In-Field Measurement of Plant Architecture Traits Using Ground-Based Platforms

Piyush Pandey and Sierra Young

Abstract

This work provides a high-level overview of system design considerations for measuring plant architecture traits in row crops using ground-based, mobile platforms. High-throughput phenotyping technologies are commonly deployed in isolated growth chambers or greenhouses; however, there is a need for field-based systems to measure large quantities of plants exposed to natural climates throughout a growing season. High-throughput methods using ground-based mobile systems collect valuable phenotypic information at higher temporal resolutions compared to manual methods (e.g., handheld calipers and measuring sticks). Additionally, the close proximity to plants when using ground-based systems compared to aerial platforms enables plant phenotyping at the organ level. While there is no single best platform for obtaining ground-based plant measurements across crop varieties with different planting configurations, there are a wide range of off-the-shelf systems and sensors that can be integrated to accommodate varying row widths, plant spacing, plant heights, and plot sizes, in addition to emerging commercially available platforms. This chapter will provide an overview of sensor types suitable for phenotyping plant size and shape, as well as provide guidance for deployment with ground-based systems, including push carts or buggies, modified tractors, and robotic platforms.

Key words Field-based phenotyping, Plant architecture, Row crops, High-throughput phenotyping

1 Introduction

High-throughput phenotyping technologies are commonly deployed in isolated growth chambers or greenhouses [1]; however, there is an emerging need for field-based systems to measure large quantities of plants exposed to natural climates throughout a growing season [2, 3]. Field-based data collection is a critical part of improving crop production as field conditions enable the expressions of the relationships between environmental and genetic factors through phenotypic variation [4–6]. Specifically, plant architecture traits, such as stem width and plant height, are highly correlated with biomass [7–9], and biomass predictions throughout the season can assist with the breeding process and enable site-

specific crop management decisions. High-throughput methods using mobile systems enable the collection of valuable phenotypic information at higher temporal and spatial resolutions compared to manual methods (e.g., handheld calipers or measuring sticks), which can accelerate the breeding and crop improvement processes.

Mobile, ground-based field systems, such as modified tractors or ground robots, are becoming increasingly popular for phenotyping applications because they enable the collection of high-resolution data at the individual plot or plant level through the combined use of high-precision GPS technology, image analysis, and geospatial analysis. While there is no “one-size-fits-all” platform for obtaining ground-based plant measurements across many crop varieties, there are a wide range of systems that can be augmented with sensors to accommodate differences in row widths, plant spacing, plant heights, and plot sizes. These ground-based platforms are especially useful if the trait of interest cannot be seen from an aerial view after the canopy closes, such as measuring stems in sorghum or brace roots in maize.

The type of mobile platform best suited for field-based data collection varies widely; however, the requirements for capturing usable, quality image data from onboard sensors are similar across platforms. Care must be taken to ensure proper sensor placement, synchronization, calibration, resolution, and collection rate. Additionally, climatic factors including sun, wind, precipitation, dust, and temperature can affect both the performance of the platform and the quality of the data collected. It is best practice to understand and plan for these challenges prior to in-field data collection. This chapter will provide an overview of technical approaches and design considerations of choosing appropriate sensors and developing a ground-based system for measuring plant architecture traits in the field.

2 Field-Based Systems: Approach and Technical Description

2.1 *Mobile Platform Form Factor*

A common approach to collecting high-throughput phenotypic data requires the use of a ground-based, mobile platform of some form to traverse the field and continuously take measurements with a variety of sensors. Modified tractor systems are a viable option if they are readily available at the field site; otherwise, they can be cost-prohibitive. Additionally, larger tractor systems may prevent the collection of under-canopy measurements due to size limitations. Portable buggies and push carts are a low-cost option that can be fabricated in house, such as the systems used in [10–12] (Fig. 1). Additionally, commercialized options exist, such as the phenoMobile® Lite buggy system developed by the Australian Plant Phenomics Facility (www.plantphenomics.org.au). However, in tall row crops, such as maize or energy sorghum, systems that

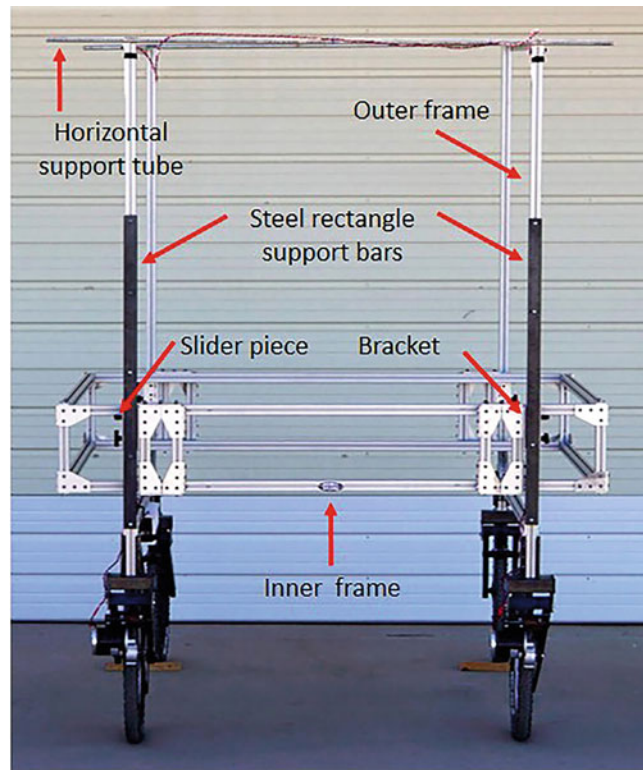


Fig. 1 “The Professor,” a mobile field cart for phenotyping applications. Originally from Thompson et al. 2018 [11]. (Licensed under CC BY 4.0 <https://creativecommons.org/licenses/by/4.0/>. No changes were made)

travel over the rows prohibit data collection late in the growing season due to chassis clearance limitations.

Small, portable robotic systems, such as the Husky (Clearpath Robotics, Waterloo, Canada, Fig. 2), or a programmable four-wheel drive robot development platform manufactured by Super-Droid Robots (Fuquay-Varina, NC, USA), can be driven between rows through the field as an alternative if chassis clearance is a major limitation and concern. Some off-the-shelf phenotyping robotic systems are now being sold with an accompanying software interface equipped with visualization from on-board cameras that can be operated manually at the field site. For example, the TerraSentia platform from EarthSense (www.earthsense.co) is a small, portable integrated data collection and robotic platform capable of capturing under-canopy data and measurements in row crops by traveling between rows and can be operated autonomously or manually. Note that platform modifications using methods of control theory, navigation, and guidance are needed to enable autonomous operation for remote-controlled systems, the methods of which are outside the scope of this chapter.



Fig. 2 “Clearpath Husky,” courtesy of Jakub Halun. (Licensed under CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0/>. No changes were made)

2.2 Vehicle Operation and Positioning

2.2.1 Mobile Ground Robots

Manual control is readily available with off-the-shelf systems and requires either a hardware controller, such as an RC transmitter, or a tablet interface with a wireless connection to the platform. Both options typically ship standard with most commercially available robots. Manual physical controllers can enable more precise control, although they may be an expensive add-on. If tablets are readily available, Wi-Fi-enabled robots can be controlled via software interfaces; however, this may not be the best option if the signal is too weak or degraded in field situations (e.g., far distances in dense canopy cover). For platforms with autonomous capabilities, positioning can be enhanced with precise GPS equipment (i.e., real-time kinematic (RTK) positioning). While convenient for operation, precise positioning systems may contribute significantly to the overall cost of the platform.

2.2.2 Modified Tractors

If tractors are available at the field site, their implements can be augmented with phenotyping sensors and driven through the field [13–15]; form factors for achieving this may include either a horizontal sensor boom mounted to the front of the tractor or a tractor-pulled cart system. The operation of a modified tractor system requires a trained operator or technician available during data collection, and, if possible, the speed of the vehicle should be set on cruise control to maintain a consistent desired speed for data collection.

2.2.3 Push Carts and Buggies

Push carts and buggy systems are operated by hand at the field site. Because human operation is highly variable, it is recommended to mount a speedometer to the cart so a consistent speed can be

achieved. Alternatively, motor encoders and a cruise control module can be added to the system to enable operation at a set speed [11]. Relative to other systems, buggies and carts require the most human effort to operate; however, they are often the most affordable systems.

2.2.4 Vehicle Speed

Desired system speed should be maintained at a steady state. Most systems operate in a range of approximately 0.1 m/s to 0.80 m/s [11, 12, 14, 16, 17], although the exact desired speed will be determined by sensor resolution, frame rate, and system capability (*see Note 3*). There is typically a “sweet spot” between vehicle speed, frame rate, and resolution to capture the required information without storing and managing unnecessarily large amounts of data. Data can always be down-sampled after field deployment if the vehicle speed was determined too slow for a specified frame rate.

2.3 Sensors for Measuring Plant Architecture

Multiple types of sensors have been used to characterize and quantify plant architecture traits. Time-of-flight (ToF) infrared depth sensors, stereo camera systems, and light detection and ranging (LIDAR) sensors are all suitable options to measure physical plant attributes and are often fused with RGB or other spectral sensors for obtaining additional phenotypes [18]. The following sections provide detail about each of these sensor types and best use practices in the context of deployment with ground vehicles for measuring plant architecture traits, including stem width, plant height, aboveground biomass, and canopy cover. Table 1 contains a list of sensor types suitable for measuring a variety of traits related to plant size and shape.

Table 1
List of plant architecture traits and sensors commonly used to measure them

Physical trait	Sensor type					
	Time-of-flight		LIDAR/laser imaging		Stereo camera systems	
	Top view	Side view	Top view	Side view	Top view	Side view
Plant height	✓ ^a	✓	✓	✓	✓	✓
Stem width		✓		✓		✓
Canopy cover	✓		✓		✓	
Leaf angle distribution	✓	✓	✓	✓	✓	✓
Volume (biomass)			✓	✓		

^aDepending on the sensor resolution

2.3.1 Time-of-Flight
(ToF) Infrared Sensors

Depth imaging using ToF cameras works by illuminating the scene with a light source and observing the reflected light. In general, the depth resolution and accuracy depend on the emitting power of the light source, and the depth accuracy of ToF sensors is normally within a range of millimeters to centimeters. Although the light entering the sensor has both an ambient and reflected component, depth information is only embedded in the reflected component; therefore, ToF sensors can be sensitive to changes in ambient lighting. It has been shown that distance error increases with the light intensity [19], so reducing the ambient lighting can improve the signal-to-noise ratio.

Raw data for ToF systems is typically in the form of a 2D grayscale depth image (Fig. 3) and can be processed using computer vision techniques to extract desired measurements [20]. Compared to other 3D laser scanning devices, ToF sensors can operate at much higher frame rates (up to 160 fps). One limitation of active depth sensors, however, is low resolution [21–23], although time-of-flight and stereo vision data create more precise depth maps when combined [24]. While ToF sensors normally have lower spatial resolutions compared to laser scanners, the next generation of continuous ToF sensors will likely have improved resolution [25]. Table 2 contains examples of available ToF depth sensors that are commercially available, with an emphasis on listing open-source options.

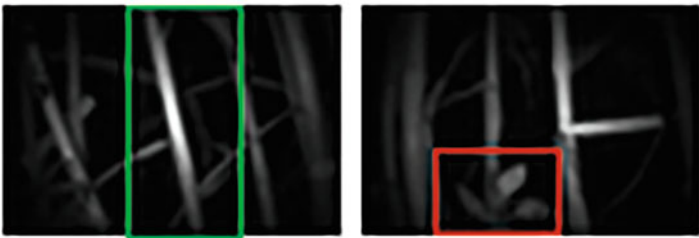


Fig. 3 An example time of flight image of sorghum stems in the field. Note that occlusion from tiller and other vegetation near the ground can occur, as the stem in the box on the left is visible, while the stem in the box on the right is not

Table 2
A list of commercially available time-of-flight sensors

Manufacturer	Model	Resolution	Frame rate	Range	Field of view
PMD Technologies	Flexx	224 × 171	Up to 45 fps	0.1–4 m	62° × 45°
PMD Technologies	Monstar	352 × 287	Up to 60 fps	0.5–6 m	100° × 85°
Lucid Vision Labs	Helios	640 × 480	Up to 60 fps	1.5–6 m	65° × 46°
Basler AG	tof640	640 × 480	Up to 20 fps	0–13 m	57° × 43°

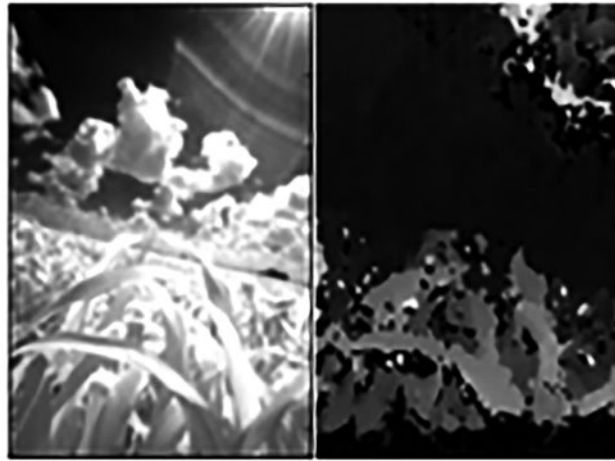


Fig. 4 Sample stereo image (left) and disparity map (right) of a sorghum canopy

For in-field phenotyping, a relatively short depth measurement range of a few meters is sufficient for deployment with ground vehicles that operate in close proximity to the plants. A side view and top view of the plant are both common configurations for phenotyping with ToF sensors [16, 19, 26]. If these sensors are deployed with a top-down view directly under the system, the image area will be generally more shaded with strong, direct sunlight being prevented from shining on the area of interest. In general, a side view of the plant is required for measuring stem width using ToF sensors, but for leaf angle distribution, plant height, and canopy cover, a top-down view is sufficient. Note that a depth resolution that is insufficient for measuring parameters of small plants and plants during early emergence may become sufficient later in the growing season as the crop growth progresses, or vice versa.

2.3.2 Stereo Camera Systems

Stereo imaging is the determination of 3D structure of a scene using two or more images of the scene, each taken from a different viewpoint. Three-dimensional information can be extracted by examining the relative positions of objects in the two images, and relative depth information is obtained in the form of a disparity map (Fig. 4). Stereo vision techniques can be implemented using a manufactured stereo camera or two independent cameras mounted in stereo that are calibrated. Examples of off-the-shelf stereo camera manufacturers and models are included in Table 3.

Camera calibration is a necessary first step to obtain intrinsic system parameters, including focal length, principal point, and radial and tangential distortion, as well as the extrinsic parameters of the stereo rig, including rotation matrix and translation vector. After calibration, the major steps of stereo imaging include the

Table 3
A list of commercially available stereo camera systems

Manufacturer	Model	Resolution	Field of view
Code Laboratories	DUO MLX	752 × 480 @ 45 fps	165° (H)
Stereolabs	ZED	3840 × 1080 @ 30 fps	90° (H) × 60° (V)
Point Grey Imaging	Bumblebee XB3	1280 × 960 @ 16 fps	66° (H)
Bosch	Stereo video camera	1290 × 960 @ 30 fps	50° (H)
Luxonis	OAK-D	1280 × 800 @ 120 fps	71.8° (H)

following [27]: (i) pre-processing, where well-defined feature characteristics are identified; (ii) feature matching, where correspondence is established between features; and (iii) 3D structure determination, which uses principles of epipolar geometry and triangulation to complete the reconstruction process. Each of these processes has been well-studied, and established algorithms exist for processing stereoscopic data [28].

Stereo camera systems, when mounted to ground-based platforms, afford the ability to measure a wide variety of physical traits. When positioned from above, stereoscopic images have been used to perform segmentation of leaf regions [29, 30] and canopy structure changes [29]. When positioned to capture a side view of the plant, stereo reconstructions can obtain leaf surface areas [31–33], plant height [16, 20, 33, 34], and stem width [34]. For field operations, it has been suggested that strong sunlight conditions and poor matching processes due to a lack of texture in the images can limit the applications of using stereo vision for phenotyping [35]; therefore, stereo cameras may yield better results when collecting data in shady conditions and positioning the camera such that there are unique features, such as contrast between plant and soil.

2.3.3 Light Detection and Ranging (LIDAR)

LIDAR is an active remote sensing system that emits light from a rapidly firing laser and records the reflected light energy that returns to the sensor to measure distances. While LIDAR was originally developed for airborne scanning, mobile laser scanner (MLS) LIDAR sensors can be used for measuring 3D properties of objects from moving vehicles. MLS LIDAR can measure three-dimensional (x, y, z) information (usually in the form of a dense point cloud, Fig. 5), obtain thousands of points per second, and measure the intensity of the reflected light. Using LIDAR as an active phenotyping sensor has two advantages, as noted by Jimenez-Berni et al. [36]: (i) it can be operated regardless of ambient lighting and (ii) it obtains direct measurements of plant architecture. Additionally, LIDAR is fairly robust in outdoor

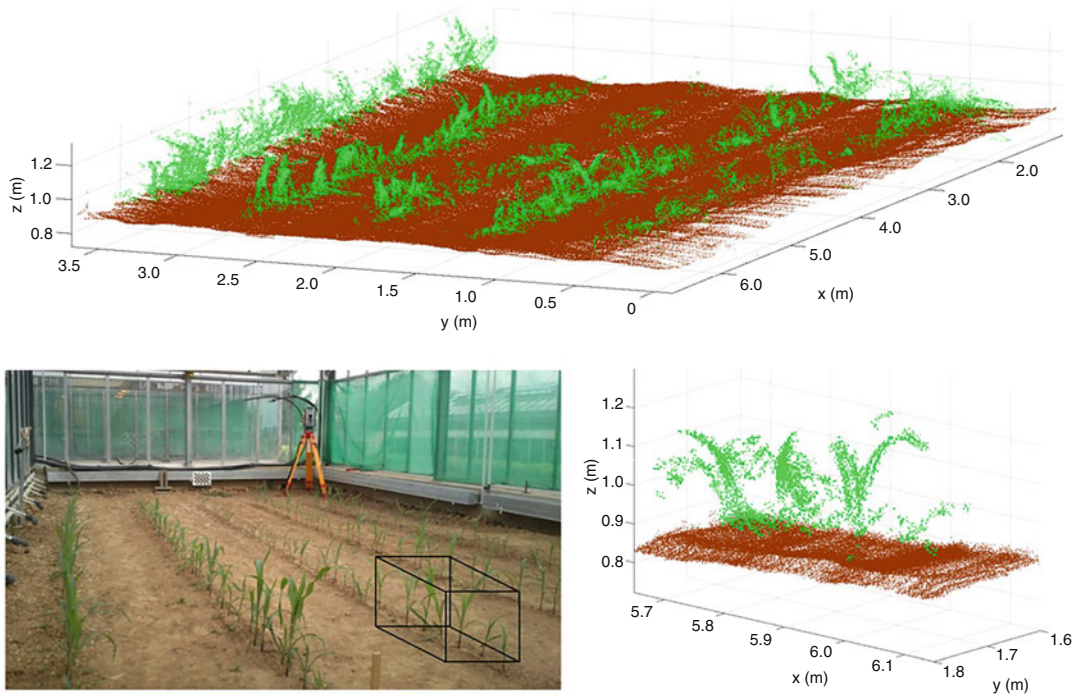


Fig. 5 Example image of LIDAR plant point cloud data, originally from Garrido et al. 2015 [47]. (Licensed under CC BY 4.0 <https://creativecommons.org/licenses/by/4.0/>. No changes were made)

conditions; however, issues may occur if wind, rain, dust, and other effects induce scanning noises [37].

While LIDAR sensors are advantageous, they generate large amounts of data, can be cost-prohibitive, and require multiple post-processing steps. Before LIDAR data can be used to calculate architecture traits, the raw LIDAR must be transformed into a point cloud; then, the point cloud must be filtered, and outliers removed. To estimate traits from the point cloud, the data can be transformed into a voxel-based format, after which plant features (e.g., leaf, stem) and traits (e.g., leaf area, height) can be extracted. LIDAR has been used successfully with ground-based phenotyping systems to obtain estimates of plant height [36, 38], biomass [36, 39], crop density [39], and canopy cover [36]. To measure canopy height, LIDAR sensors can be placed either above or below the canopy, so long as a fraction of the pulses penetrates the canopy. Note that LIDAR data can provide direct estimates of biomass, while stereo camera systems and active infrared depth sensors measure traits used in crop models to estimate aboveground biomass.

2.4 Sensor Control and Data Acquisition

For phenotyping applications, it is critical that sensor data are properly georeferenced so images can be referenced to the appropriate plot for extracting traits during post-processing workflows;

additionally, the data acquisition parameters (frame rate, exposure, etc.) must be controlled individually for each sensor. There are many system architecture and control approaches that achieve synchronized sensor operation, and an example system is outlined below that attempts to reduce the number of required components and utilize commercially available sensors that operate via standard interfaces, such as Ethernet or USB. The components of this example data acquisition system and their functions are described below.

2.4.1 On-Board Computer System and Server

An on-board computer system can be used to control the data acquisition and synchronization via software and receive GPS information over serial connection, and there are a wide range of suitable systems for these tasks. The NUC Mini PCs (Intel Corporation, Santa Clara, CA, USA) are powerful systems with customizable memory, storage, and operating system, all at a convenient size ($11.7'' \times 10.2'' \times 3.9''$) and weight (around 500 g) for deployment; however, these systems are not ruggedized and also require headless operation or additional displays. Ruggedized laptops, such as the Latitude Rugged series (Dell, Round Rock, TX, USA), are sealed from sand and dust, although they require larger mounting space. All computer systems will typically require AC power supplies on-board the platform.

2.4.2 Image Capture

It is advantageous to choose camera sensors that are open source or that operate via standard USB/serial connections, which allow for the creation of specific software programs or tools that can customize multiple data capture parameters (sometimes through sensor-specific software development kits [SDKs]). Additionally, programs can be written so that data acquisition is event triggered or begins on system startup, which is useful when the computer system (e.g., NUC or Raspberry Pi) requires headless startup and operation.

2.4.3 Sensor Synchronization

The on-board server can be used to synchronize the data streams in time using the system's internal clock. All data should be time-stamped and GPS-tagged to enable plot-level mapping of phenotypic traits during post-processing workflows. The GPS system used on-board should provide the current time and a pulse per second (PPS) signal that can be used for timing other devices. Using the GPS as a time source can be more accurate than relying on the system's NTP clock, especially when the system is not connected to a network in the field. The GPS time can be used to update the system's NTP server and can also be used in the sensor log files.

2.4.4 Data Storage

Image data can accumulate storage space quickly; therefore, it is often necessary to utilize an external storage device, such as a solid-state drive (SSD), where the data are immediately saved, organized, and stored during the field trials (*see Note 6*). Also, using portable,

external storage devices facilitates manual data transfer from the on-board system to a server for post-processing, such as a networked desktop machine or workstation.

3 Materials and Equipment

3.1 *Phenotyping System Hardware*

When choosing hardware for a phenotyping experiment, there are many variables to consider that impact both the platform and sensors, such as the crop, desired traits, resolution of the data, cost, availability, and required expertise of operation. The lists of materials provided in this section are intended to be a generic set of minimum equipment required to collect phenotypic data using a combination of stereo, ToF, and/or LIDAR sensors in the field, from a ground-based platform.

1. Mobile, ground-based deployment system (tractor, robot, push cart/buggy).
2. Sensor mounting hardware.
3. On-board power supply.
4. DC to AC converter.
5. On-board computer and mounting platform.
6. GPS device with appropriate connection (i.e., serial, Ethernet).
7. Phenotyping sensors with appropriate interface connector.
8. External storage device.
9. Optional: actively cooled electronics chamber.

A note about power supplies: A power load estimation can be conducted for the sensor system by summing the predicted average current draw for each subsystem. If the platform is an off-the-shelf robotic system or a modified tractor, the power supply for the mechanical drive system will likely already be integrated and will be separate from the sensor and data acquisition power supply. Sensor and computer system power usage can be estimated by using programs developed for this purpose or by using a real-time Watt meter during testing operations. Depending on power consumption and desired length of operation, lead acid batteries are a low-cost power supply option; however, they can be heavy and have a large form factor. Alternatively, lithium-ion polymer (LiPo) batteries can be used, which offer much higher capacities in smaller form factors, although they require extra care for safe operation, charging, and storage.

3.2 *Field Deployment Items*

During initial development, additional items can be brought to the field for rapid prototyping and data quality assurance. This list of items and their suggested uses are given for best practices and can

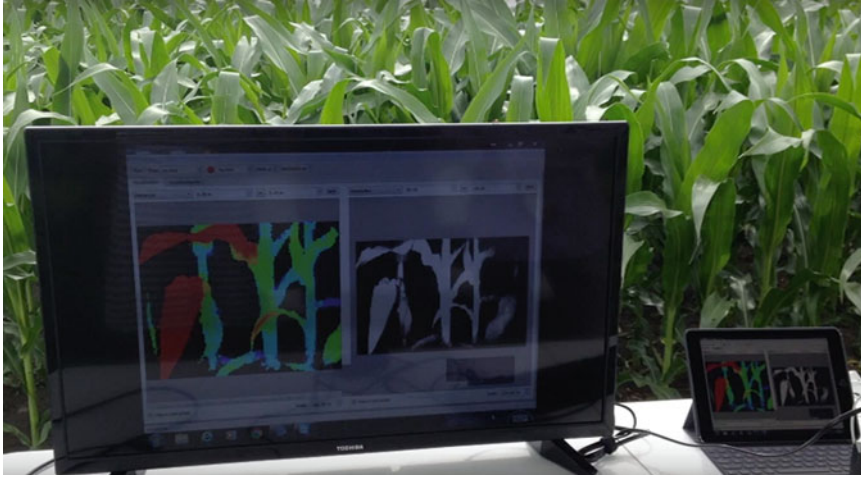


Fig. 6 Screens showing sensor data streaming in real time from a phenotyping system in the field. During the prototyping stage, viewing the data in real time can assist with determining optimal sensor placement

serve as a checklist. The items below are also useful for field demonstrations and field days.

1. *Video displays*: Using a display in the field during initial testing is important for data quality control to ensure correct positioning and capture (Fig. 6). These monitors can connect to the on-board system over a wireless network through a screen sharing software, such as TeamViewer for Windows or Remmina for Linux and Unix-based systems (both are free to use).
2. *Wireless network hotspot*: If the area of interest does not have any wireless connection (which is often the case), it is useful to establish a local, ad hoc network for establishing headless control of the on-board computer. A straightforward way to do this is by using a device such as the Jetpack® MiFi® 8800L (Verizon Communications, New York, NY, USA) or a personal cellular device to set up a local wireless hotspot using the 4G LTE network. Note that the device needs to stay within connectivity range of the phenotyping system.
3. *Pop-up tent*: This is critical for safety reasons and to preserve ground station and platform electronics when performing field work in extreme heat or conditions with unexpected heavy precipitation (Fig. 7).
4. *Tables and chairs*: These assist in safety and comfort when conducting field work. They are not essential, but are highly suggested, especially for field demonstrations.



Fig. 7 Temporary tent, tables, and chairs set up in the field for a demonstration. This tent enclosure protected the display electronics from heat and unexpected precipitation

5. *Portable generator*: A power supply such as a generator can be used to power the visual monitor displays, as well as supply backup power should the on-board power fail during critical data collection.
6. *Extension cables*: Extension cables can be used to bring power from the generator to the required devices in the tent and to the system in the field during critical failure.

4 Field Deployment Methods

This section outlines general steps required for setting up and deploying a phenotyping system using the example above, and includes items to consider during deployment.

4.1 Data Acquisition Setup (Pre-deployment)

1. Calibrate the selected camera or sensor. This can be done using manufacturer-specified procedures or standard functions available in the OpenCV library [40] depending on the sensor selected.
2. Configure the computer system NTP clock to use the GPS PPS signal.
3. Develop a custom program to operate the cameras as needed – this is typically done using the sensor SDK, Python or C++ with

OpenCV, LabView, and/or JavaScript. The exact functions for adjusting sensor parameters will vary depending on the manufacturer capabilities and software, but at a minimum the following should be addressed:

- Capture frame rate
 - Exposure (static or auto)
 - Image timestamp
 - Image GPS stamp, using the system's NTP clock
 - IMU measurements (especially if using LIDAR)
 - Filename
 - Output directory/file location
 - Image metadata
4. Create a startup script (e.g., bash shell or batch) that will run the data acquisition script(s) from **step 2** on machine startup if using a headless system (optional).

4.2 Hardware Setup and Deployment

1. Securely mount the following to the mobile system:
 - On-board power supply and converters
 - Computer system and/or server
 - Phenotyping sensors and their connection cables
2. Make sure the cables are secured to the frame of the system, so they do not snag during deployment.
3. Connect your computer/processor and sensors to the power supply.
4. Optional: if operating in regions that get extremely hot, an actively cooled chamber may be required to keep the electronics at proper operating temperatures, for example, by mounting a fan in an air-permeable enclosure or forced air conditioners.
5. Turn on the system and ensure all sensors are operational.
6. Begin moving the system through the field, either by a remote controller or manually. Remember to maintain a stable position and steady speed (use cruise control, if possible; otherwise, monitor speed using speedometer).
7. Continue collecting data throughout the deployment area of interest. Remember that all data will be time and GPS stamped, so identifying data at the plot level can occur during post-processing.
8. Optional: Aim for field deployment times early in the morning to avoid extreme heat and direct sunlight conditions, especially if infrared sensors are used (*see Notes 1 and 2*). However, if also collecting spectral data from other reflectance sensors, desired deployment times may be different.

9. Depending on the experimental setup, develop a map of the area needed to measure. If using a system with sensors mounted above the crop, one pass may be needed to image an entire plot; if using a system that operates between the rows, multiple passes may be needed for full plot coverage [although one set of image data per plot may suffice [16]].
10. If required, collect ground truth data for each trait using a “gold standard” method (*see* **Notes 4** and **5**).
11. When data collection of the desired area is completed, properly shut down all electronics. Transfer the data to a central or permanent server. Remove the data from the on-board storage device to ensure capacity for future deployments.

5 Broad Overview of the Post-processing Workflow

After data collection, the post-processing workflow begins in which there are a set of required, general steps for extracting the traits of interest. These steps are outlined in Fig. 8. First, the data must be georeferenced, and if needed, 3D reconstructions made. Then, the data are reduced by extracting data at the plot level, after which outliers and noise can be removed. Additional modifications can also be made to image data to enhance the segmentation and trait extraction procedures, such as applying a filter or modifying contrast.

Before the traits of interest can be measured, the data must be segmented to separate the plant from the rest of the image or data set. Segmentation methods can be separated into two extremely broad categories: (i) region-based approaches, which find sets of pixels with corresponding sets of properties, and (ii) contour-based approaches, which include edge detection and continuity analyses [41]. Both supervised and unsupervised methods exist within each of these categories. For a review of computer vision approaches for phenotyping, see [42].

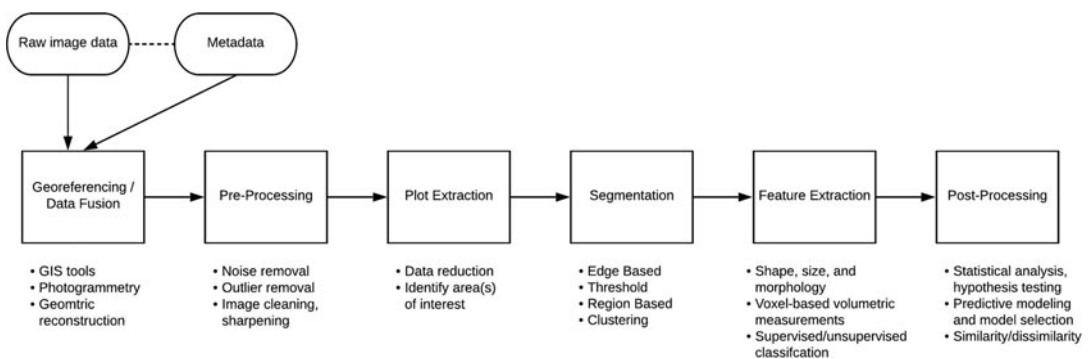


Fig. 8 A general overview of the required post-processing workflow steps for phenotypic data. The boxes represent the major steps, while the text underneath includes examples of more specific approaches and sub-processes for each step

After image segmentation, plant architecture traits such as width, area, and shape can be measured. To obtain direct measurements, methods such as simple pixel or voxel counting are useful. For more advanced shape analysis, a set of tools is available in PlantCV [43]; additionally, a list of available plant image analysis software (including PlantCV) is available at <https://www.plant-image-analysis.org/>.

6 Summary

This chapter provided an overview the technical approach and design considerations of choosing appropriate sensors and developing a ground-based system for measuring plant size and shape in the field. Time-of-flight, stereo, and LIDAR sensors have all been widely used to measure various plant architecture traits, including stem width, plant height, canopy cover, leaf angle distribution, and biomass estimations, among others. To deploy these sensors, ground-based systems such as modified tractors, robotic systems, carts, and buggies offer advantages in data resolution for phenotyping plants at the organ level. Regardless of platform, care must be taken to ensure proper sensor placement, synchronization, calibration, resolution, and collection rate. Understanding best practices and anticipating challenges that might arise during data collection are important to ensure the collection of consistent, reliable phenotypic data. Field-based data collection is a critical part of improving crop production, and high-throughput methods using ground-based systems enable the collection of valuable phenotypic information to accelerate the plant improvement process.

7 Notes

1. To reduce ambient lighting, an enclosure can be fabricated around the sensors to effectively block out sunlight [14, 44]. If ambient light is needed for imaging (e.g., if RGB sensors are also used), the enclosure can be artificially lighted to specified, optimal conditions.
2. It is not recommended, even if the system is ruggedized, to operate in weather conditions with strong winds, precipitation, or strong sunlight. These conditions have the potential to adversely affect both the operation of the system and the quality of data collected by introducing noise.
3. When determining vehicle speed and frame rate, we have found that from experience this can be achieved by trial and error and will depend on the sensor specifications, platform capabilities, and exact phenotyping needs. For example, traveling at a speed



Fig. 9 A sorghum stem, illustrating the differences in width between different locations along the stem

of approximately 0.3 m/s with a sensor at 15 fps generated more than enough data for extracting traits at the plant level using a ToF sensor with a side view of the plant [16].

4. It is important that ground truth data are collected for enough plants for each trait of interest during the initial deployment. These data should be used for validation purposes during post-processing and can be collected at the individual plant or plot level, so long as it is possible to compare the sensor-made measurements to the ground truth measurements.
5. Develop a standard for measurement of each trait for consistency between the sensor-derived measurements and the manually collected ground truth measurements. For example, when measuring stem width of sorghum, the width varies depending on proximity to the internode (Fig. 9); additionally, the stem has an elliptical shape. Therefore, depending on orientation and distance from the internode, the ground truth measurement may vary significantly from the measurement extracted from the image data. To remedy this, develop standards for measurement that are consistent between ground truth data and sensor-derived data. For example, stem width could be measured perpendicular to the field midway between the first and second internode.

6. Ongoing efforts are being made to standardize the collection and storage of phenotypic observations to enable interoperability between data providers [45]. MIAPPE, or Minimum Information About a Plant Phenotyping Experiment, is a checklist of metadata required to adequately describe a plant phenotyping experiment and a software to validate, store, and disseminate data [46]. It is recommended to follow these working standards and the complementary implementation tools that support its application, especially if the data will be shared or made available. For more information, go to www.miappe.org/.

References

1. Batz J, Méndez-Dorado AM, Thomasson AJ (2016) Imaging for high-throughput phenotyping in energy sorghum. *J Imaging* 2(1). <https://doi.org/10.3390/jimaging2010004>
2. Fiorani F, Tuberosa R (2013) Future scenarios for plant phenotyping. *Ann Rev Plant Biol* 64: 267–291
3. Furbank R (2009) Plant phenomics: from gene to form and function. *Funct Plant Biol* 36: 10–11
4. Sankaran S, Khot LR, Espinoza CZ et al (2015) Low-altitude, high-resolution aerial imaging systems for row and field crop phenotyping: a review. *Eur J Agron* 70:112–123
5. White JW, Andrade-Sanchez P, Gore MA et al (2012) Field-based phenomics for plant genetics research. *Field Crops Res* 133:101–112
6. Araus JL, Cairns JE (2014) Field high-throughput phenotyping: the new crop breeding frontier. *Trends Plant Sci* 19(1):52–61
7. Wight JP, Hons FM, Storlien JO et al (2012) Management effects on bioenergy sorghum growth, yield and nutrient uptake. *Biomass Bioeng* 46:593–604
8. Casa AM, Pressoir G, Brown PJ et al (2008) Community resources and strategies for association mapping in sorghum. *Crop Sci* 48:30–40
9. Ehlert D, Horn H-J, Adamek R (2008) Measuring crop biomass density by laser triangulation. *Comp Electron Agric* 61(2):117–125. <https://doi.org/10.1016/j.compag.2007.09.013>
10. Berni JAJ, Zarco-Tejada PJ, Suarez L et al (2009) Thermal and narrowband multispectral remote sensing for vegetation monitoring from an unmanned aerial vehicle. *IEEE Trans Geosci Remote Sens* 47(3):722–738. <https://doi.org/10.1109/TGRS.2008.2010457>
11. Thompson AL, Conrad A, Conley MM et al (2018) Professor: a motorized field-based phenotyping cart. *HardwareX* 4:e00025. <https://doi.org/10.1016/j.ohx.2018.e00025>
12. Conley JWM (2013) A flexible, low-cost cart for proximal sensing. *Crop Sci* 53:1646–1649
13. Andrade-Sanchez P, Gore MA, Heun JT et al (2014) Development and evaluation of a field-based high-throughput phenotyping platform. *Funct Plant Biol* 41:68–79
14. Busemeyer L, Mentrup D, Möller K et al (2013) BreedVision - a multi-sensor platform for non-destructive field-based phenotyping in plant breeding. *Sensors* 13(3):2830
15. Bao Y, Tang L, Breitzman MW et al (2019) Field-based robotic phenotyping of sorghum plant architecture using stereo vision. *J Field Robot* 36(2):397–415
16. Young SN, Kayacan E, Peschel JM (2018) Design and field evaluation of a ground robot for high-throughput phenotyping of energy sorghum. *Precision Agric*. <https://doi.org/10.1007/s11119-018-9601-6>
17. Fan Z, Sun N, Qiu Q et al (2021) A high-throughput phenotyping robot for measuring stalk diameters of maize crops. 2021 IEEE 11th annual international conference on CYBER technology in automation, control, and intelligent systems (CYBER), 128–133
18. Atefi A, Ge Y, Pitla S, Schnable J (2021) Robotic technologies for high-throughput plant phenotyping: contemporary reviews and future perspectives. *Front Plant Sci* 12:611940
19. Klose R, Penlington J, Ruckelshausen A (2009) Usability study of 3D time-of-flight cameras for automatic plant phenotyping. *Bornimer Agrartechnische Berichte* 69(12):93–105
20. Baharav T, Bariya M, Zakhora A (2017) Computing height and width of *in situ* sorghum plants using 2.5D infrared images. In: IS&T international symposium on electronic imaging

- computational imaging XV, Burlingame, CA, USA, pp 1–14
21. Pinto AM, Costa P, Moreira AP et al (2015) Evaluation of depth sensors for robotic applications. In: 2015 IEEE international conference on autonomous robot systems and competitions, 8–10 April 2015, pp 139–143. <https://doi.org/10.1109/ICARSC.2015.24>
 22. Day B, Bethel C, Murphy R et al (2008) A depth sensing display for bomb disposal robots. In: 2008 IEEE international workshop on safety, security and rescue robotics, Oct. 21–24 2008, pp 146–151. <https://doi.org/10.1109/SSRR.2008.4745892>
 23. Foix S, Alenya G, Torras C (2011) Lock-in time-of-flight (ToF) cameras: a survey. *IEEE Sensors J* 11(9):1917–1926. <https://doi.org/10.1109/JSEN.2010.2101060>
 24. Jiejie Z, Liang W, Ruigang Y et al (2008) Fusion of time-of-flight depth and stereo for high accuracy depth maps. In: 2008 IEEE conference on computer vision and pattern recognition, 23–28 June 2008, pp 1–8. <https://doi.org/10.1109/CVPR.2008.4587761>
 25. Li F, Chen H, Yeh C-K et al (2019) High spatial resolution time-of-flight imaging. In: *Computational Imaging III*, 2019. International Society for Optics and Photonics, p 1066908. <https://doi.org/10.1117/12.2303794>
 26. Chéné Y, Rousseau D, Lucidarme P et al (2012) On the use of depth camera for 3D phenotyping of entire plants. *Comp Elect Agric* 82:122–127
 27. Dhond UR, Aggarwal JK (1989) Structure from stereo—a review. *IEEE Trans Syst Man Cybern* 19(6):1489–1510
 28. Hartley R, Zisserman A (2003) *Multiple view geometry in computer vision*. Cambridge University Press, Cambridge, UK
 29. Biskup B, Scharr H, Schurr U et al (2019) A stereo imaging system for measuring structural parameters of plant canopies - BISKUP - 2007. *Plant Cell Environ* 30(10):1299–1308. <https://doi.org/10.1111/j.1365-3040.2007.01702.x>
 30. Nguyen TT, Slaughter DC, Maloof JN et al (2016) Plant phenotyping using multi-view stereo vision with structured lights. In: *SPIE commercial + scientific sensing and imaging*, 17 may 2016 2016. Autonomous air and ground sensing systems for agricultural optimization and phenotyping. SPIE, p 9
 31. Klodt M, Herzog K, Töpfer R et al (2015) Field phenotyping of grapevine growth using dense stereo reconstruction. *BMC Bioinform* 16(1):143. <https://doi.org/10.1186/s12859-015-0560-x>
 32. Song Y, Glasbey CA, van der Heijden GWAM et al (2011) Combining stereo and time-of-flight images with application to automatic plant phenotyping. In: Heyden A, Kahl F (eds) *Image analysis*. Springer, Berlin Heidelberg, pp 467–478
 33. Tuong Nguyen TC, Slaughter D, et al (2016) Comparison of structure-from-motion and stereo vision techniques for full in-field 3D reconstruction and phenotyping of plants: An investigation in sunflower. Paper presented at the 2016 ASABE Annual International Meeting, Orlando, Florida, Jul 17–20, 2016
 34. Salas Fernandez MG, Bao Y et al (2017) A high-throughput, field-based phenotyping technology for tall biomass crops. *Plant Physiol* 174(4):2008–2022
 35. Li L, Zhang Q, Huang D (2014) A review of imaging techniques for plant phenotyping. *Sensors* 14(11):10.3390/s141120078
 36. Jimenez-Berni JA, Deery DM, Rozas-Larraondo P et al (2018) High throughput determination of plant height, ground cover, and above-ground biomass in wheat with LiDAR. *Front Plant Sci* 9(237). <https://doi.org/10.3389/fpls.2018.00237>
 37. Lin Y (2015) LiDAR: an important tool for next-generation phenotyping technology of high potential for plant phenomics? *Comp Electron Agric* 119:61–73
 38. Tilly N, Hoffmeister D, Cao Q et al (2014) Multitemporal crop surface models: accurate plant height measurement and biomass estimation with terrestrial laser scanning in paddy rice. *Remote Sens* 8(1):23. <https://doi.org/10.1117/1.JRS.8.083671>
 39. Saey W, Lenaerts B, Craessaerts G et al (2009) Estimation of the crop density of small grains using LiDAR sensors. *Biosyst Eng* 102(1):22–30. <https://doi.org/10.1016/j.biosystemseng.2008.10.003>
 40. The Open Source Computer Vision (OpenCV) Library. <https://opencv.org/>
 41. Malik J, Belongie S, Leung T et al (2001) Contour and texture analysis for image segmentation. *Internat J Comp Vision* 43(1):7–27. <https://doi.org/10.1023/A:1011174803800>
 42. Li Z, Guo R, Li M et al (2020) A review of computer vision technologies for plant phenotyping. *Comp Electron Agric* 176:105672
 43. Gehan MA, Fahlgren N, Abbasi A et al (2017) PlantCV v2: image analysis software for high-throughput plant phenotyping. *PeerJ* 5. <https://doi.org/10.7717/peerj.4088>

44. Jiang Y, Li C, Robertson JS et al (2018) GPhe-
noVision: a ground mobile system with multi-
modal imaging for field-based high throughput
phenotyping of cotton. *Sci Rep*
8(1):1213–1213. [https://doi.org/10.1038/
s41598-018-19142-2](https://doi.org/10.1038/s41598-018-19142-2)
45. Ćwiek H, Krajewski P, Klukas C et al (2015)
Towards recommendations for metadata and
data handling in plant phenotyping. *J Exp Bot*
66(18):5417–5427. [https://doi.org/10.
1093/jxb/erv271](https://doi.org/10.1093/jxb/erv271)
46. Ćwiek-Kupczyńska H, Altmann T, Arend D
et al (2016) Measures for interoperability of
phenotypic data: minimum information
requirements and formatting. *Plant Methods*
12(1):44. [https://doi.org/10.1186/s13007-
016-0144-4](https://doi.org/10.1186/s13007-016-0144-4)
47. Garrido M, Paraforos SD, Reiser D et al (2015)
3D maize plant reconstruction based on geor-
eferenced overlapping LiDAR point clouds.
Remote Sens 7(12):10.3390/rs71215870



Chapter 16

Design and Construction of Unmanned Ground Vehicles for Sub-canopy Plant Phenotyping

Adam Stager, Herbert G. Tanner, and Erin Sparks

Abstract

Unmanned ground vehicles can capture a sub-canopy perspective for plant phenotyping, but their design and construction can be a challenge for scientists unfamiliar with robotics. Here we describe the necessary components and provide guidelines for designing and constructing an autonomous ground robot that can be used for plant phenotyping.

Key words Robotics, Phenotyping, Design, Construction, Sub-canopy

1 Introduction

Advances in genetics research have revolutionized the agricultural industry in ways that have vastly improved crop yield and resistance; however, progress has slowed due to the challenges of identifying desirable genetic traits. Some beneficial traits are identifiable by physically measuring features of plants and can be measured non-destructively using portable handheld tools. While manual field measurement is possible, given that there can be thousands of plants per acre, and factoring in labor costs and variability between human crop scouts, consistent data acquisition at the scale necessary to integrate phenotyping into genetic pipelines can be difficult to obtain. Fully and semi-autonomous robotic systems offer a solution for gathering vast amounts of data in the field to relieve the phenotyping bottleneck [1].

The goal of this manuscript is to offer a guide, in layman's terms, on how to design and construct a ground mobile robot. It targets an audience of scientists or practitioners with little or no prior knowledge in robotics. We begin with a general discussion on component selection, to guide design choices for systems tailored for sub-canopy data collection. Then, a detailed example outlines the construction of a tracked robot platform (see Fig. 1) intended



Fig. 1 Adjusting cameras on a modified SuperDroid LT2 Tracked ATR platform. For simplicity, the design described in this method includes a single camera sensor, but it is easy to add more sensing capabilities. Pictured here are additional sensors, including one extra RGB camera, an RGBD camera, analog video transmission, GPS, and external waterproof electronics housing. Camera rails allow quick and easy outfitting of the robot for various phenotyping experiments

for sub-canopy terrain conditions in corn fields at the University of Delaware.

2 Design Considerations

Whether a robot is to be built from scratch or purchased off the shelf, there are many factors to consider when choosing the right equipment. By combining our own experience with that found in literature, we bring together important design considerations for ground mobile robots. We step through each component of a sub-canopy robot and provide some practical insights to streamline the design process.

2.1 Application Constraints

The scientist's crop and traits of interest determine the robot's specifications and set quantitative robotic design metrics (i.e., width must be less than 50.8 cm; sensor height must be greater than 243.84 cm) that help a robot designer compare tradeoffs between features of the robot. For example, the amount of money one can allocate to the development of a robot might put a constraint on cost. Because sub-canopy robots operate in a particularly confined environment, some of the important metrics relevant to sub-canopy phenotyping are discussed.

Minimum row spacing is one field characteristic that has an important impact on the size of the robot. The width of the robot is typically constrained to the planting row width minus two times the maximum stalk diameter (*see Note 1*). Ideally, a robot should be as narrow as possible; a thin design can accommodate better motion uncertainty and reduce the chances of collision with plants. There are tradeoffs, however: decreased width will reduce stability in the roll axis (tipping to the left and right when driving forward down the row). If the terrain is expected to be very flat and dry, then the tip-over risk is less significant.

Crop height and intended robot and/or plant localization accuracy are factors that affect the robot's height and sensing capabilities. In general, it is best to keep the robot's center of mass low (*see Note 2*) in order to increase stability; however, sub-canopy systems may be required to measure features at the plant's extrema where the top canopy can exceed heights of 4.5 m (e.g., bioenergy sorghum [2]). If measurements are required at a prescribed height along the stalk or information on average crop height is to be measured, a lightweight mast can provide a solution. This mast can double as a mount for an accurate global positioning sensor (GPS) known as a real-time kinematic (RTK) GPS, but coupled with rough or uneven terrain, a mast can also cause tipping and positional errors and may interact with the top of the plants as shown in Fig. 2.

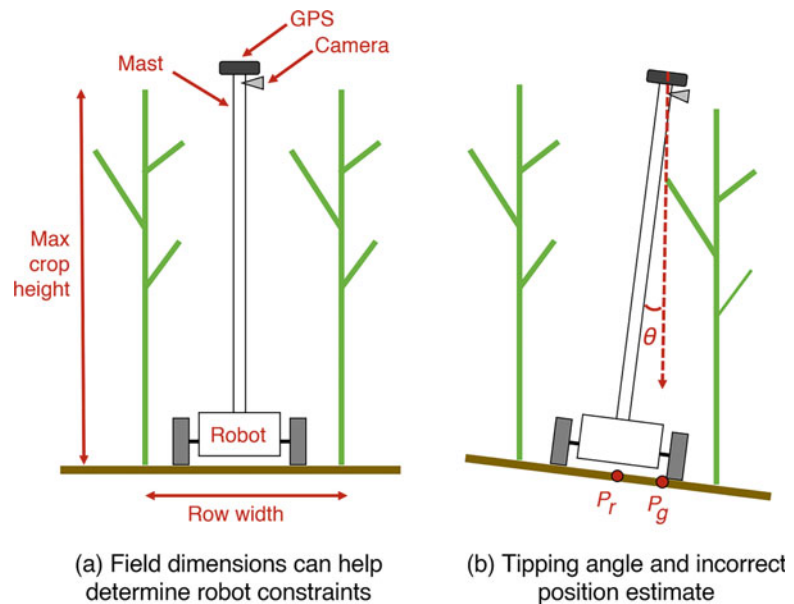


Fig. 2 (a) Row width and maximum estimated crop height are important constraints for sub-canopy robot platforms. (b) Tipping angle θ , caused by uneven ground or rough terrain, can cause instability of the robot and lead to a false position estimate P_g due to a resulting offset of the GPS from the robot's actual position P_r

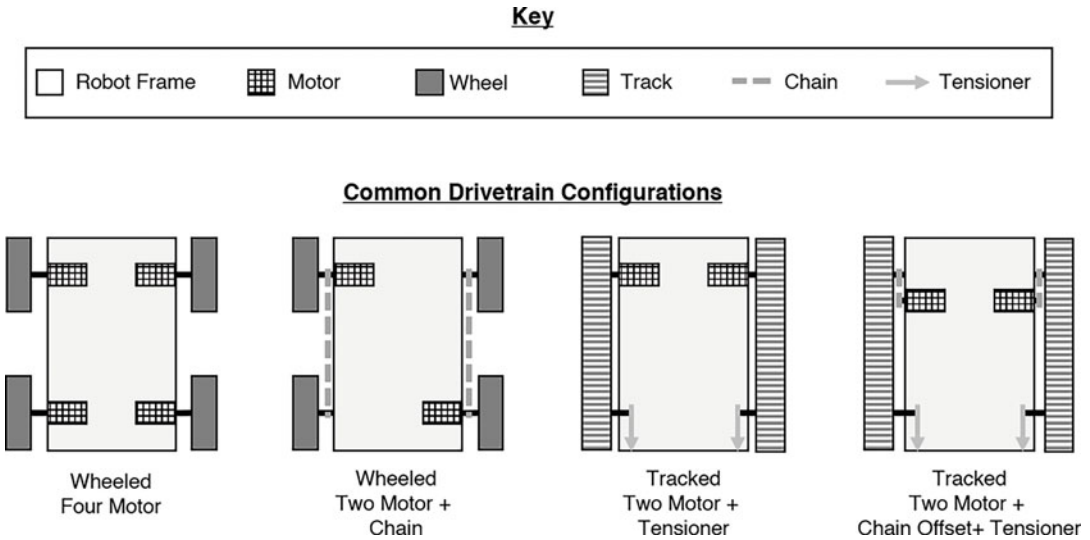


Fig. 3 Examples of the most common configurations for ground mobile robots used for sub-canopy phenotyping. The arrows indicate a tensioner that moves in the direction of the arrowhead when it is tightened

2.2 Frame

After setting the restrictions on the robot’s physical dimensions and design configuration, the mechanical structure of the robot is considered. Some off-the-shelf designs offer pre-built frames (e.g., SuperDroid Robots, Inc.) that can be easily modified [3]. Generally, it is more practical to keep the frame lightweight because weight impacts overall runtime; in addition, lighter robots are easier to transport. A heavy robot can also leave depressions on the ground and is more likely to get stuck in mud. Cost often limits the use of exotic lightweight materials such as composites and titanium. Painted and galvanized steel and aluminum are common choices because they offer good structural rigidity at a relatively low cost. Although steel is heavy and more difficult to machine, it is easy to weld, and its low cost makes it a good choice for prototype designs. Aluminum is a convenient, lightweight material that can be bent into complex shapes, machined at low cost, while resisting corrosion. Some designs offer modularity by using a combination of clamps and tubes [4].

2.3 Drivetrain

The drivetrain is the system that allows the robot to get around in its environment. It can be configured in many different ways [5] and typically includes as major components wheels or tracks, motor (of different size and placement), motor drivers and controllers, and batteries (*see* Fig. 3). Sub-canopy robots are most often designed with either tracks or wheels. Tracks tend to cause less rutting and are less likely to get stuck in rough terrain, but they require more power, are more expensive, typically require more complicated transmission mechanisms, and can pick up weeds or tillers. Wheeled setups are often four-wheel drive, with one side

running independent of the other (differential drive), allowing the robot to turn in place. Wheels, however, are susceptible to high centering –when the center of the robot gets stuck on rough terrain – but their low cost and ease of use make them a good option when the terrain is not particularly challenging.

Motors come in several forms: there are brushed, brushless, stepper, and servo motors, but the easiest to control and most commonly used for ground mobile robots are brushed direct current (DC) motors. Other types of motors require special sensing and/or timing control to make them spin, whereas brushed motors contain physical brushes (hence their name) that connect electrically their spinning and stationary parts. The speed and torque requirements are two specifications that guide the selection of motor size. Knowing how much field coverage is desired is a good place to start because it can narrow motor and gearing combinations to a desired number of revolutions per minute (RPM). For example, the TERRA-MEPP sub-canopy platform used two 24VDC DG-158A wheelchair motors with a 135 RPM no-load output speed to cover 0.4 hectares in 2 h [6]. Average motor torque is difficult to compute a priori during the design process because calculations rely on complex ground interactions (depending on if it is muddy, dry, level of weeds, or tiller interaction) and friction between transmission components; however, it is possible to calculate upper bounds that help with motor selection.

Once the brushed DC motors are selected, they should be paired with a brushed DC motor driver or motor controller. These electrical components allow the robot's on-board central processing unit (CPU) to control the amount of power sent to the motors. A motor controller is typically more expensive than a motor driver, because it has additional functionality such as speed and position control (when encoder feedback is provided) or switching between remote control (RC) and autonomous control modes. The specification sheet associated with the motors will give information on “amperage at stall torque” which is an indication of how much the motor driver/controller should be able to output in the worst (most challenging) case. If a stall condition is encountered in the field, an undersized motor driver/controller can overheat causing thermal shutdown or damage to the circuitry (*see Note 3*).

Battery selection bridges the gap between drivetrain and electrical subsystem because it influences the entire system. Ultimately, the size of the battery, described in Amp-hours (Ah), should be determined based on the estimated power draw from the motors plus the power draw of all sensors and the CPU. Most of these values can be found on specification sheets, but the motor output is heavily dependent on sizing, losses in power transmission, and environmental conditions. Similarly, with motor sizing, a bound

on maximum and minimum runtimes of the robot and average working conditions can help establish an average battery power consumption, helping to determining proper battery capacity.

2.4 Electrical

It is unlikely that a robot will run on a common, consistent voltage, as different components may require different input voltages. Once the robot's function and capabilities are well established, a designer may assemble a wiring harness with carefully measured lengths and connectors that are mounted with zip ties or cable clips along the frame of the robot. During the prototyping phase however, it is useful to have flexibility so that new sensors can be tested without redesigning the entire electrical system. It is useful to include a power and ground distribution block, which is a physical component connecting wires to a single source, typically using convenient screw terminals. Voltage regulators can help adjust the voltage depending on the requirements of the motors, sensors, and CPU. Depending on expected operating conditions, moisture can be an important consideration affecting how to weather-proof the electrical components. A completely sealed box integrated into, or attached to, the frame can provide a waterproof setup, but operating such a system in high-temperature environments may cause overheating of the electronics. Off-the-shelf CPU coolers are available with incorporated air filters to prevent dust from entering the compartment and can be mounted to an electrical box to maintain safe operating temperatures (*see Note 4*).

2.5 On-Board Computer (CPU)

An on-board computer is necessary if a designer prefers to do any computation on the robot. Selection of an on-board computer depends heavily on both the desired level of autonomy and number of sensors that will be connected to the system. For cases where very little processing is required, a microcontroller such as an Arduino, which can only run a single script, may suffice. For navigation and real-time processing, however, a microprocessor is needed. The processing power of microprocessors found on single board computers, such as the Raspberry Pi 3, are generally limited but may still be suitable for applications with minimal to moderate autonomy or sensing. Fully autonomous or semi-autonomous ground robots would likely require a mini PC. The Intel NUC is one commonly used mini PC because of its highly compact form factor and range of available options for processing power. Scientists who require vast amounts of sensor data for processing feedback in real time have also networked multiple dedicated computers serving different subsystems to avoid processing conflicts between subsystems [7].

For navigation, some off-the-shelf autopilots such as Pixhawk or the Navio2 can be combined with freely available software to streamline the problem of navigating the robot along a path, but are limited to common robot configurations, and they are not as

flexible for path planning as a custom setup. For systems where geolocation is not required (i.e., when tags in the field identify plants) or if the ground platform is to be teleoperated, then the on-board computer and sensing system can be decoupled from the motor control.

2.6 Navigation Sensors

Data from several sensors can be fused together to get accurate position and orientation of the robot as it travels through the field. If precision planting coordinates are available, then paths through the rows can be established; otherwise, the robot can be driven manually while capturing GPS data to identify waypoints. If the robot is to navigate without relying on some assumed plant arrangement and/or centimeter-level accuracy is needed while following a predetermined path, then the accuracy of a real-time kinematic (RTK) GPS is indispensable. For sub-canopy robots, the GPS is typically placed on top of a mast in order to obtain a good signal; the canopy will otherwise obstruct the connection and make GPS unreliable. When the GPS unit is attached to a vertical mast, then errors in position due to the roll and pitch of the robot can be significant and naturally increase with mast length. An accelerometer (*see Note 5*) on the base can mitigate the problem by providing additional data for tilt correction. In conditions where wheel slippage is minimal, encoders can be used to get reasonable estimates of position, although one has to keep in mind that skid-steer vehicles rely on wheel slippage during turns, and that introduces errors in estimates of orientation. A gyroscope (*see Note 6*) coupled with a magnetometer (*see Note 7*) can be used in tandem to resolve unambiguously the platform's orientation. Fusing sensor data in a Kalman filter (*see Note 8*), a standard robot position/velocity estimation algorithm, can help obtain better position estimates; Kalman filter implementations are available in several open-source software packages. Redundancy in sensor data becomes increasingly important as the robot's speed increases, especially considering the relatively slow 5–10 Hz data rate from GPS, and can make position estimation more robust to occasional sensor outliers and failures. Using visual odometry (*see Note 9*), a state-of-the-art method of camera feeds with other sensor measurements and for which implementation software is also freely available, is another promising strategy for localizing a robot; it does require, however, relatively more processing power and careful calibration and can be prone to errors when used in highly dynamic environments.

2.7 Phenotyping Sensors

There are many imaging modalities available for collecting data relevant to phenotyping; however, for in situ field sensing, we consider only non-destructive samplers, cameras, and sensors (*see Note 10*) that can be readily mounted on a ground mobile robot.

To help with the daunting task of selecting between the wide range of available options, this section briefly describes some of the most common sensor types and how they can be used for phenotyping.

2.7.1 RGB Camera

Arguably the most common sensing modality, single-sensor cameras capturing visible light (VIS) can be used with no geographic data if some label is present on a plant to associate it to its images [8]. Images can be taken at high resolution at the expense of a larger file size. Many features can be identified from VIS data, especially in structured environments when software tools such as PlantCV (<http://plantcv.danforthcenter.or>) are available [9]. However, differences in lighting and occlusions in the field may require additional processing for data analysis. Recently, machine learning algorithms have been used to extract features from plants by first training such algorithms with datasets labeled by human experts [10]. Some common features extracted include leaf area index (LAI), plant height, stem thickness, yield estimates, and leaf and stand count [11, 12].

2.7.2 Spectral Camera

Spectral cameras are designed to pick up light from individual spectra not visible to the human eye. By focusing light emission as a result of excitation by specific wavelengths of light, scientists can study internal characteristics of a plant non-destructively and before they become apparent by sensing visible light. The Normalized Difference Vegetation Index (NDVI) and Photochemical Reflective Index (PRI) are two classical indices obtained by spectral imaging and used to quantify plant health [13, 14]. There are two categories of spectral cameras – hyperspectral and multi-spectral. Hyperspectral cameras are orders of magnitude higher resolution than multi-spectral cameras and are a recent focus of research. These cameras can capture thousands of narrow bandwidths compared to multi-spectral cameras which acquire data from 5 to 12 much wider spectral bands; however, the data from hyperspectral cameras can be overwhelming, and the requirement for careful calibration makes these new cameras impractical for some applications [15].

2.7.3 Stereo Camera

By combining two cameras with a known separation distance, stereo cameras use correspondence between images to calculate distances in the form of disparity maps and provide estimates of depth for objects in the image. Accurate distance estimates rely on matching features between images and perform poorly in low light where features can be difficult to distinguish [16].

2.7.4 Time-of-Flight (ToF) Sensor

By emitting infrared (IR) light and measuring the return time of the reflected light on a low-resolution camera, distances can be estimated directly. These sensors are generally low resolution compared to RGB cameras but can offer depth without the computation of stereo camera setups. ToF cameras work well in

low light, but are susceptible to noise in direct sunlight because the sun's IR emission can conflict with the sensor's output IR. Despite their relatively low resolution and sensitivity to ambient light, ToF are generally favorable for determining features in outdoor agricultural settings [17].

2.7.5 RGBD Camera

By combining RGB with a ToF camera, RGBD cameras give a similar output as stereo cameras, but require less computation. Similar to ToF sensors, RGBD can be sensitive to lighting conditions, and they have limited range due to the necessary matching between depth and RGB cameras. The Microsoft Kinect is an example of a low-cost RGBD camera that is commonly incorporated into robotics projects.

2.7.6 LIDAR

Light detection and ranging (LIDAR) uses a pulsed laser to measure distances at very high resolution with minimal noise compared to stereo and ToF camera sensors. Although these sensors can capture features in very fine detail, they require accurate knowledge of the position of the sensor and are significantly more expensive than other remote sensing methods. Sensor feedback is expressed in dense point clouds and can generate massive amounts of data that can be challenging to abstract to useful features [18].

2.8 Communication

Bandwidth and signal attenuation are the most important considerations when deciding on how to communicate between the robot and a ground station (a stationary computer where the human operator tracks robot progress). While an autonomous robot is capable of navigating on its own, keeping human operators in the loop safeguards against unanticipated challenges. For remotely controlled systems, communication delays (latency) should be reduced because a lag in video or control command can result in crop damage.

Generally, commands are sent over radio frequencies; the range is exceptional, and real-time control commands usually require very low bandwidth. Wi-Fi is another way for communication with field robots offering higher bandwidth and allowing for video transmission, but at the expense of decreased range. For systems operating in the range of cellular towers, LTE can also be used to transmit high bandwidth data. Although cellular LTE offers superior range, it can have fluctuating latency and is generally too slow for real-time teleoperation. Analog video has been gaining popularity for remote controlled vehicles, driving costs down and increasing reliability for these components. Although the analog video feed is not directly suitable for processing, it is a good option for teleoperation because it provides a long-range option with low latency and very little setup compared to other methods.

2.9 Software

Depending on the level of autonomy required, it can be relatively quick to get a robot running in the field. Sensors usually come with software for operating them on Windows, Mac, and Linux making it possible to collect data out of the box. At a minimum, a robot can be set up with teleoperated navigation and an RGB camera with on-board data storage. Although this requires intervention by a human operator, it can provide useful data from the hard-to-reach sub-canopy region and is a good place to start. If plants are tagged with visible identifiers, then a human can quickly assign images to different plants, and computer vision algorithms can achieve measurements of stalk thickness, leaf area index (LAI), canopy density, and other useful characteristics.

3 Construction of a Tracked Robot Platform

To provide a concrete example, we describe next the construction of a basic phenotyping robot. Off-the-shelf components are used wherever possible to make this platform reproducible with very little robotics background and at relatively low cost. Specifically, the base is a SuperDroid LT2 Tracked ATR package including the frame, drivetrain, power, and electrical components (*see Note 11*).

3.1 Materials

- Frame (SuperDroid LT2 Tracked ATR robot platform).
- Drivetrain (tracks + tensioner + chain offset).
- Power and electrical (24VDC lead acid, RoboteQ MDC2460 2x60A 60 V motor controller, RC controller, RC receiver).
- On-board computer (Raspberry Pi 3 with 16GB SD micro card).
- Phenotyping and navigation sensors (camera and camera rails).
- Communication (Wi-Fi and radio).
- Software (Linux operating system).
- Ground control computer (Linux operating system).
- Misc parts (mounting hardware, foam weather proofing tape, 5 lb mounting tape, USB AB cable, RS232 connector).

3.2 Tools

- Phillips screwdriver.
- Adjustable wrench.
- 1/8" Allen wrench.
- 8 mm Allen wrench.
- Chain breaker.
- Measuring tape.
- 1/4" box end wrench.

- Wire cutter.
- Crimping tool.
- Power drill.
- 3/8" drill bit.
- 15/32" drill bit.
- HDMI cable.
- Monitor.
- Keyboard.
- Mouse.

3.3 Mechanical Assembly

The mechanical assembly of a tracked robot requires accurate spacing between track wheels and a mechanism for tensioning the belt. Driving the wheels directly from the output of the motor's gearbox is not recommended and can lead to pre-mature failure of expensive motor components. The schematic (Fig. 4) shows a visual representation of the most important components throughout the robot assembly, but purchase of a SuperDroid robot kit comes with an instruction manual including images. Here we summarize the steps to build the LT2 Tracked ATR robot – all components listed are included in the SuperDroid robot kit except for the camera rails and clamps which were purchased from a camera accessory company, SmallRig.

1. Using a Phillips screwdriver and adjustable wrench, bolt front axles (shafts for mounting the track wheels) to aluminum frame with eight #10-32 bolts and nuts.
2. Mount two drive motors, IG52-04 24VDC 285 RPM Gear Motors, using a Phillips screwdriver to tighten four M5 machine screws per motor.
3. Using the adjustable wrench, loosely mount track tensioning blocks using #6 hardware.
4. Slide rear axle into the tensioning blocks (mounting points that will be used to tension the tracks), and tighten two lock collars using 1/8" Allen wrench to prevent the axle from sliding left or right.
5. Slide #25 sprockets over motor shafts, but do not tighten their set screws until **Step 11**.
6. Install back-handle strap using Phillips screwdriver to tighten four tensioning screws.
7. Assemble two drive wheels with sprockets and two idler wheels using a Phillips screwdriver and ten #10 screws per wheel.

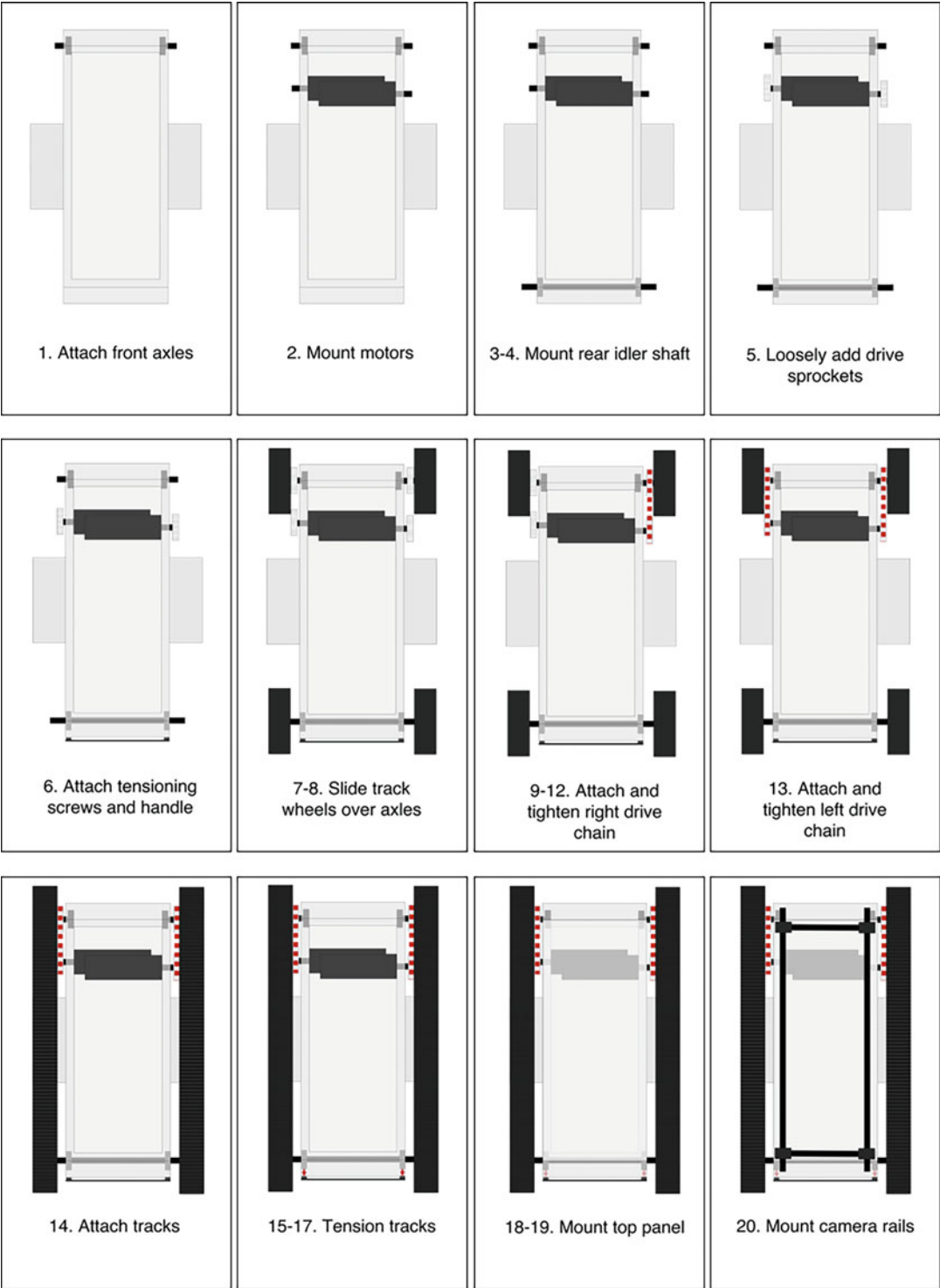


Fig. 4 A visual guide representing the mechanical assembly of a SuperDroid LT2 Tracked platform

8. Slide wheel with sprocket on the front axle using a thrust bearing as a spacer on the inner and outer faces of the wheel followed by a lock collar tightened using a 1/8" Allen wrench.
9. Slide the motor forward toward the front wheel, and then wrap the #25 chain around both sprockets.
10. Using a chain breaker, remove excess chain and connect chain master link.
11. Slide motor sprocket along the motor shaft so it is aligned with the wheel sprocket, and then tighten motor sprocket set screw with a 1/8" Allen wrench.
12. Pull the motor away from the wheel until the chain is tight, and then tighten motor mounting screws using Phillips screwdriver.
13. Repeat **Steps 8–12** for the other side.
14. Slide the tension blocks all the way forward, and then roll 2.75 inch molded spliceless tracks over the front and rear wheel on each side.
15. Tighten the four tension screws with the Phillips screwdriver at the rear of the robot evenly until the tracks are tight.
16. On each side, using a measuring tape, measure from the front axle center to the rear axle center to make sure both sides have equal separation. If the measurement on the right side is larger than the left, then use the Phillips screwdriver to tighten the tension screws on the left side until they are equal and vice versa.
17. Insert a 1/4" box end wrench between the robot frame and the track to tighten the four screws that hold each tension block to the robot frame.
18. Cut 1/4" wide strips from the foam weather proofing tape, peel off the adhesive backing, and attach the tape around the entire perimeter of the top panel. This is used to cover the opening in the top of the frame.
19. Mount top panel using 16 Phillips head screws using Phillips screwdriver.
20. Drill four 3/8" holes 1" × 1" from the corners of the top mounting plate, and then mount small rig camera rail clamps using an 8 mm Allen wrench (*see Note 12*). The rail clamps allow the mounting rails to be adjusted depending on the application.

3.4 Electrical Assembly

The electrical assembly describes the connection of batteries, motor drivers, power regulators, and sensors. Here we use two 12VDC lead-acid batteries in series to create a 24VDC power source. This is fed directly to the motor driver and regulated down to 5VDC for a

Raspberry Pi 3 (on-board computer) and 12VDC to power an on-board Wi-Fi router. All components in this section are included in the LT2 Tracked ATR package except for power and ground distribution blocks, 5VDC regulator, 12VDC regulator, Raspberry Pi 3, and Wi-Fi router. The regulators chosen should provide enough Amps to power the electronics connected to them. A more powerful Wi-Fi router will reduce latency, and its signal can be improved by adding a long-range antenna on the ground control station, robot, or both. The camera is powered directly from a USB port on the Raspberry Pi 3. IG52-04 24VDC 285-RPM gear motors are connected to a RoboteQ MDC2460 2x60A 60 V motor controller. Make sure to consider locations for mounting the electronics and routing the wiring – these aspects are often overlooked (*see* **Note 13**).

1. Install two 12 V 8 Ah sealed lead-acid batteries, one on each side inside the aluminum LT2 chassis.
2. Slide threaded rods through the battery mounting tabs of the aluminum frame, and tighten nuts over each end with an adjustable wrench (*see* **Note 14**).
3. Use mounting tape or screws to attached power and ground distribution blocks inside the frame.
4. Drill a 15/32" hole at the rear of the robot (in black plastic switch plate), and slide the power switch through the hole, and then tighten the nut, securing the switch using an adjustable wrench (*see* **Note 15**).
5. Use mounting tape or screws to mount power regulators (5VDC, 12VDC) inside the frame.
6. Use mounting tape or screws to mount RoboteQ MDC2460 2x60A 60 V motor controller inside the frame.
7. Connect the RoboteQ motor controller to USB port of the Raspberry Pi 3 using a USB AB cable.
8. Use RS232 connector to connect RC control to the RoboteQ.
9. Use camera rail clamp to mount camera in desired location on camera rails, and then feed camera's USB cable back to a USB port on the Raspberry Pi 3.
10. Confirm connections with the schematic in Fig. 5 before powering on the robot.

3.5 Software Setup

Software is an important part of a robot and can be increasingly complex depending on the level of autonomy desired. On the other hand, without adding software, a teleoperated robot can still collect useful data. A barebone system is possible where a GoPro or other standalone camera can be attached to the robot as it is driven through the field. Images can then be extracted afterward by retrieving them from the camera's memory card. This method is

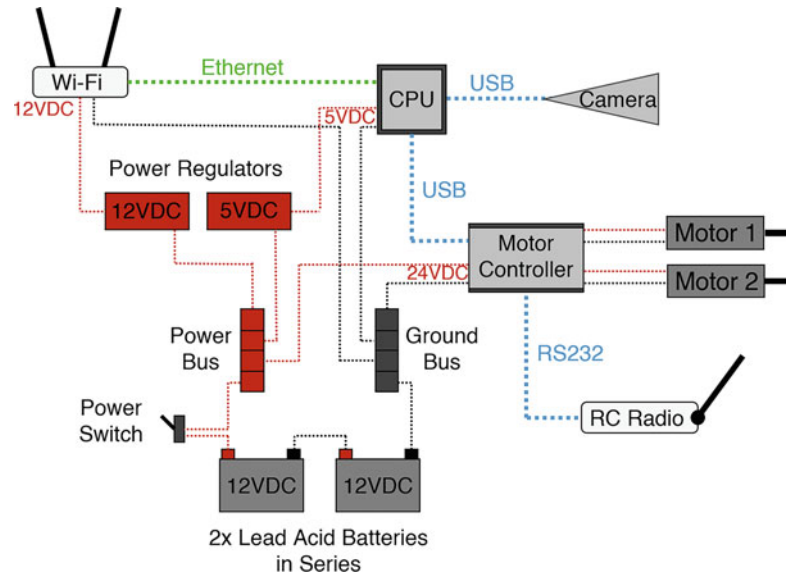


Fig. 5 Schematic representation of electrical connections

very limited because the robot has no control over the camera or its data. Instead, we offer a middle-ground solution by describing a setup that can be easily expanded upon. A single camera on-board the robot streams video over Wi-Fi to the ground control station where the user can collect images in real time while also using the camera to navigate the robot using RC control. Additional cameras or sensors can be added, and the control can be automated.

The robot uses the open-source Robot Operating System (ROS) on a Linux Ubuntu Mate operating system and sends data to a ground control computer running Linux Ubuntu 16.04 LTS. We assume Linux is installed on both ground control station and on-board computers (*see Note 16*). ROS is a set of software utilities and libraries that is convenient because it allows for easy implementation of new sensors and stores/handles sensor data in a convenient organized structure. It's worth mentioning that these steps can be challenging to approach with no prior experience with Linux, Raspberry Pi 3, or ROS. We hope that the steps provided will help to guide an inexperienced reader and have provided links where necessary to help reinforce more complex steps.

1. Startup (boot) Linux (Ubuntu Mate distribution) by powering on the Raspberry Pi 3 (*see Note 17*).
2. Install ROS Kinetic using the command line (*see Note 18*).
 - (a) Boot Raspberry Pi 3 with Ubuntu Mate installed.
 - (b) Connect Wi-Fi to the internet, open a web browser, and navigate to <http://wiki.ros.org/kinetic/Installation/Ubuntu> where line-by-line commands can be copy/pasted into the terminal (where command line text can be run).

- (c) Highlight each command, and right-click on the computer mouse to select copy, and then right-click in the terminal window, and select paste.
 - (d) Press Enter to run each command.
 - (e) Once complete, test the installation by first navigating to <http://wiki.ros.org/ROS/Tutorials/InstallingandConfiguringROSEnvironment> in the web browser.
 - (f) Create a ROS workspace by copy/pasting terminal commands as in **Steps 2.c–2.d**.
3. Download the usb_cam ROS node (http://wiki.ros.org/usb_cam) which is a collection of programs and camera utilities packaged with the camera driver and various parameters that make it easy to use. Calibration is one file that should be updated depending on the camera that is used. In this step, it is easiest to plug in an HDMI cable, mouse, and keyboard to confirm the camera works directly on the Raspberry Pi 3.
 - (a) Once the ROS environment is created, open a new terminal.
 - (b) Install the usb_cam ROS package by typing “sudo apt-get install ros-kinetic-usb-cam” followed by Enter.
 - (c) Once installed, plug in the USB camera, and open two separate terminal windows. In the first window, run “roscore.” In the second terminal window, run “roslaunch usb_cam usb_cam_node.” This will publish the raw data from the camera to the computer.
 - (d) To view the camera data, open one more terminal window, and run “rqt_image_view.” From the pull-down in the top left of the popup window, select “/usb_cam/image_raw” to confirm the camera output.
 4. Setup remote camera triggering and video streaming over Wi-Fi.
 - (a) Plug the Raspberry Pi 3 with Ubuntu Mate and ROS into the robot’s Wi-Fi router using an Ethernet cable.
 - (b) Boot the Raspberry Pi 3 while connected to HDMI, and type “ifconfig” into a new terminal. Find and take note of the “inet addr” which will start with 192.168.1.x where x is a unique identifier designated to the Raspberry Pi 3. Now the HDMI, keyboard, and mouse can be removed from the Raspberry Pi 3.
 - (c) On the ground control computer, log into the robot’s Wi-Fi network.

- (d) On the ground control computer, open a new terminal and run “ifconfig” to identify the ground station’s “inet addr.” It will start with 192.168.1.y where y is a unique identifier designated to the ground control computer.
 - (e) Log into the Raspberry Pi 3 remotely using secure shell (SSH) (*see* **Note 19**).
5. Stream live video from Raspberry Pi 3.
- (a) Complete **Step 4** making sure to note the unique addresses 192.168.1.x and 192.168.1.y where x and y are numbers unique to the Raspberry Pi 3 and ground control computer, respectively.
 - (b) Boot the Raspberry Pi 3 and power the robot’s Wi-Fi router (*see* **Note 20**).
 - (c) Open three terminal windows on the ground control station’s computer.
 - (d) In the first terminal, log into the Raspberry Pi 3 remotely as in **Step 4.e** (*see* **Note 21**).
 - (e) In the second terminal, run “roscore.”
 - (f) Return to the first terminal and run “roslaunch usb_cam usb_cam_node.”
 - (g) In the third terminal, run “rqt_image_view.” From the pull-down in the top left of the popup window, select “/usb_cam/image_raw.” The robot’s video feed should be displayed.
 - (h) Save images by clicking on the save file icon to the far right of the pull-down selection. This is not a fast way to collect images from the robot but will give a feel for what it is like to teleoperate a system and get data remotely. It will help with understanding the challenges of latency and can be extended to autonomously capture images or include multiple camera.

4 Notes

1. It may be more relevant to consider an estimated maximum stalk (base) diameter because branches, tillers, or brace roots extending into the row can disturb the robot’s motion if they are not considered.
2. It is best to keep heavier components closer to the ground.
3. For four-motor configurations, the 4WD operation will be severely limited if current on each side of the robot is split between a single motor driver. In this case, the designated motor input will travel the path of least resistance, and

insufficient traction for either wheel on a particular side will result in wheel spinning and loss of forward motion. It is best to allocate either one motor driver per motor or one dual motor driver per side to get true 4WD traction.

4. It is practical to pair the motor voltage with the battery output because motor output will be a heavy drain on the battery capacity. If the voltages do not match, then a voltage regulator can change the battery output at the expense of a loss in efficiency during the voltage conversion. This would get you less motor output for the same battery charge, and voltage regulators can also limit the output current to the motors.
5. Typically, accelerometers measure acceleration along their local x, y, and z axes. By measuring the force of gravity, an accelerometer can help inform the robot which direction is down. With additional computation, it can also help estimate the robot's position.
6. A gyroscope (digital) is used to determine orientation in roll, pitch, and yaw.
7. A magnetometer measures magnetic forces and acts as a compass for the robot. Magnetic fields from the motors can affect magnetometer readings so they should not be placed close together.
8. A Kalman filter combines measurements from multiple sensors to reduce uncertainty.
9. Visual odometry is a method of determining position and orientation based on analyzing changes between camera images.
10. Cameras are also a type of sensor, but we typically see the output in the form of an image. The raw data is similar to other sensors. For example, an array of 10×10 individual photo (light)-absorbing semiconductors would be a 100-pixel camera sensor. Digital values for red, green, and blue are captured for each pixel and then processed into the color image typically viewed.
11. The LT2 Tracked ATR package can also be purchased pre-assembled.
12. Four SmallRig 15mm camera rails, four SmallRig quick release clamps, and four SmallRig 90-degree rod clamps are needed to mount the small rig camera rail.
13. We have found it useful to use 5 lb outdoor mounting tape to mount electronics semi-permanently for when aspects of the robot are still in development. This prevents unnecessarily drilling extra holes in the robot's frame and saves time, but electronics should be mounted more permanently once the robot is complete. Alternatively, a mounting panel can be

made with many holes drilled into it for mounting small parts. This way the mounting panel can be replaced when components change and as a result less holes are required in the robot's frame. The goal is to keep wiring neat and to keep wires short if possible. Zip ties can be helpful for securing wires along the frame. We leave the selected method of mounting the electronics up to the reader.

14. These rods act as battery tie downs to make sure the batteries don't shift into the CPU or other electronics.
15. A fuse that prevents too much current from flowing through the system can be added but is not required.
16. These operating system (OS) procedures are well documented and outside the scope of this work.
17. In order to install Ubuntu Mate on a Raspberry Pi 3, a 16 GB (minimum) SD micro memory card is needed. If it must be installed, then download Ubuntu Mate for free (<https://ubuntu-mate.org/download/>), and format the SD card; then using free Win32DiskImager (<https://sourceforge.net/projects/win32diskimager/>), the .img file for Ubuntu Mate can be added to the SD card. Insert the SD into the Raspberry Pi 3, and follow the prompts to set up Ubuntu Mate. If the ground control station's computer does not have Linux, then a "bootable USB" can be created using a .iso image of Linux distribution Ubuntu 16.04 LTS. Booting the computer from the USB will guide setup alongside windows or as a standalone operating system. These Linux distributions were chosen because they have been shown to cause the fewest difficulties in installation and use, but they can be switched with newer versions as long as they support ROS.
18. The command line is a text interface where you can launch commands (run programs) by typing specific text into the interface window, called the terminal. You can open a terminal by pressing CTRL+ATL+T.
19. SSH is a way of sending commands to a computer on the same network. Knowing the login name and password to the Raspberry Pi 3, log into it from the ground control station computer by opening a new terminal and running the command "ssh login@address," where "login" is the login name of the Raspberry Pi 3 and "address" is the inet address found by running "ifconfig" in **Step 4.b**. A prompt will request the password from the Raspberry Pi 3. Now this terminal window on the ground control computer can be used to run programs on the robot remotely.

20. If the Raspberry Pi 3 and Wi-Fi router are connected to the robot's power source, then flipping the power switch on the robot will power both on. Otherwise, if they are not yet installed in the robot, power them independently.
21. If the Raspberry Pi 3's inet address is not recognized, then make sure the ground control station's computer is connected to the robot's Wi-Fi network. Sometimes when it is powered off, it will default to a previously saved network and lose connection with the robot.

Acknowledgments

This work is supported by the Delaware Biosciences Center for Advanced Technology Entrepreneurial Proof of Concept Grant.

References

1. Furbank RT, Tester M (2011) Phenomics - technologies to relieve the phenotyping bottleneck. *Trends Plant Sci* 16(12):635–644. <https://doi.org/10.1016/j.tplants.2011.09.005>
2. Baharav T, Bariya M, Zakhori A (2017) In situ height and width estimation of sorghum plants from 2.5D infrared images. *Elect Imaging* 17: 122–135. <https://doi.org/10.2352/issn.2470-1173.2017.17.coimg-435>
3. Zhang J, Singh A, Lofquist A et al (2018) A novel multirobot system for plant phenotyping. *Robotics* 7(4):61. <https://doi.org/10.3390/robotics7040061>
4. Grimstad L, From P (2017) The Thorvald II agricultural robotic system. *Robotics* 6(4):24
5. Bawden O, Ball D, Kulk J et al (2014) A lightweight, modular robotic vehicle for the sustainable intensification of agriculture. *Proc of 16th ARAA ACRA*, 1–9
6. Young SN, Kayacan E, Peschel JM (2018) Design and field evaluation of a ground robot for high-throughput phenotyping of energy sorghum. *Precision Ag*. <https://doi.org/10.1007/s11119-018-9601-6>
7. Mueller-Sim T, Jenkins M, Abel J et al (2017) The Robotanist: a ground-based agricultural robot for high-throughput crop phenotyping. *ICRA*:3634–3639
8. <https://doi.org/10.1109/ICRA.2017.7989418>
9. Fahlgren N, Gehan MA, Baxter I (2015) Lights, camera, action: high-throughput plant phenotyping is ready for a close-up. *Current Opin Plant Biol* 24:93–99. <https://doi.org/10.1016/j.pbi.2015.02.006>
10. Fahlgren N, Feldman M, Gehan MA et al (2015) A versatile phenotyping system and analytics platform reveals diverse temporal responses to water availability in *Setaria*. *Mol Plant* 8(10):1520–1535. <https://doi.org/10.1016/j.molp.2015.06.005>
11. Singh A, Ganapathysubramanian B, Singh AK et al (2016) Machine learning for high-throughput stress phenotyping in plants. *Trends Plant Sci* 21(2):110–124. <https://doi.org/10.1016/j.tplants.2015.10.015>
12. Li L, Zhang Q, Huang D (2014) A review of imaging techniques for plant phenotyping. *Sensors* 14(11):20078–20111. <https://doi.org/10.3390/s141120078>
13. Mulla DJ (2013) Twenty five years of remote sensing in precision agriculture: key advances and remaining knowledge gaps. *Biosyst Eng* 114(4):358–371. <https://doi.org/10.1016/j.biosystemseng.2012.08.009>
14. Gamon JA, Peñuelas J, Field CB (1992) A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. *Remote Sens Environ* 41(1):35–44. [https://doi.org/10.1016/0034-4257\(92\)90059-S](https://doi.org/10.1016/0034-4257(92)90059-S)
15. Tucker CJ (1979) Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens Environ* 8(2):127–150. [https://doi.org/10.1016/0034-4257\(79\)90013-0](https://doi.org/10.1016/0034-4257(79)90013-0)
16. Adão T, Hruška J, Pádua L et al (2017) Hyperspectral imaging: a review on UAV-based sensors, data processing and applications for

- agriculture and forestry. *Remote Sens* 9(11): 1110
17. Kise M, Zhang Q, Rovira Más F (2005) A stereovision-based crop row detection method for tractor-automated guidance. *Biosyst Eng* 90(4):357–367. <https://doi.org/10.1016/j.biosystemseng.2004.12.008>
18. Klose R, Penlington J, Ruckelshausen A (2011) Usability of 3D time-of-flight cameras for automatic plant phenotyping. *Bornimer Agrartechnische Berichte* 69:93–105



Nighttime Chlorophyll Fluorescence Imaging of Dark-Adapted Plants Using a Robotic Field Phenotyping Platform

Maria Newcomb and Nadia Shakoor

Abstract

Photosynthetic efficiency is increasingly recognized as an integration of plant responses to dynamic environments, establishing the need for data sets from both field trials and controlled environments. A robotic field scanner phenotyping platform at the University of Arizona is equipped with a high-throughput chlorophyll fluorescence imaging system capable of collecting data on field trials for genetic studies of a photosynthetic trait (F_v/F_m). A description of the fluorescence imaging system is provided in addition to methods for measurements across experimental field plots and a test to determine the impact of variable plant heights. The overall focus is on aspects of field applications of a chlorophyll fluorescence imaging system that differ from analogous systems in controlled environments.

Key words Field scanner, PSII, Chlorophyll fluorescence, Quantum efficiency, High-throughput phenotyping, Field phenotyping

1 Introduction

Chlorophyll fluorescence imaging has been widely applied as a non-contact plant phenotyping method for characterizing the activity of photosynthetic systems in indoor controlled environments [1]. Photosynthetic activity is a fundamental determinant of plant health and productivity and is the result of complex genetics. This provides opportunities for phenotype to genotype studies to guide breeding for improved crop varieties with enhanced photosynthetic traits [2]. There is increasing awareness of the importance of studying photosynthetic parameters in field environments and associated interest in chlorophyll fluorescence imaging systems that can be used as a tool in field phenotyping programs. While there are multiple benefits to controlled environment studies that can isolate particular variables of interest, the fact that photosynthesis efficiency is an integration of plant responses to dynamic

environments establishes the need for data sets from both field trials and controlled environments [3].

To date, studies using in-field automated phenotyping systems to assess photosynthesis in fluctuating light conditions have been rare [4]. Over half a decade ago, Murchie and Lawson (2013) predicted the incorporation of remote fluorescence measurements into automated high-throughput crop phenotyping methods and techniques. Here we describe a novel field imaging system and methods for chlorophyll fluorescence measurements of dark-adapted plants using the robotic field scanner system at the University of Arizona Maricopa Agricultural Center. The field fluorescence imaging system has been utilized in 2 different field trials to measure the maximum quantum efficiency of photosystem II (estimated as F_v/F_m) of dark-adapted plants for a genotyped bi-parental sorghum mapping population of 160 RILs planted in 2 replicates (320 plots) and a genotyped durum wheat diversity panel of 223 lines in 2 replicates and 2 irrigation treatments plus repeating checks (924 plots). These phenotyping scans are part of the first applications of a robotic semi-automated field phenotyping platform to obtain high-throughput chlorophyll fluorescence image data for genetic studies of a photosynthetic trait (F_v/F_m).

Quantification of chlorophyll *a* fluorescence is a useful indicator of the status of photosynthetic processes since emitted photons inform the relative rates of photochemical and non-photochemical quenching. Several reviews describe the process by which light energy absorbed by plant photosystems is dissipated as either photochemistry, heat, or re-emitted photons (fluorescence) [5–7]. The maximum quantum yield of the photosystem II (PSII) system can be determined by measuring fluorescence at known states: fully dark-adapted when PSII reaction centers are electrically neutral or open (F_0 minimal fluorescence) and following a saturating illumination when reaction centers cannot use the excitation energy as a result of charge separation and are closed (F_m maximum fluorescence) [7]. Together the fluorescence measurements at these two states can be used to determine the quantum efficiency of PSII photochemistry using the equation $F_v/F_m = (F_m - F_0)/F_m$ [6]. F_v/F_m values for plant leaves in the absence of any stress are close to 0.83 across species and indicate plant stress and overall metabolism that directly or indirectly reflects photosynthetic functions [4, 6, 8].

Incorporating chlorophyll fluorescence image data in high-throughput phenotyping programs in field environments has tremendous value and also challenges. In field studies, imaging platforms allow chlorophyll fluorescence measurements at the scale of whole plants and plant canopies and capture the variability among plants as well as within plants under natural conditions [9]. A complete system includes an illuminating light with sufficient intensity to saturate the photosystems of plant leaves over the

imaged area, a synchronized and robust camera fit with a bandpass filter, cooling capacity to accommodate the light source, and importantly a mechanism for moving the system across a field trial. The design and installation require engineering expertise and knowledge of the biophysics of fluorescence and photosynthesis parameters [9]. While commercial options for automated field imaging systems are currently limited, for example, PhenoVation [10] and Qubit Systems, it is likely that available options will increase in the near future.

The field scanner at the University of Arizona Maricopa Agricultural Center is a large robotic gantry system constructed and employed by the TERRA-REF project funded by the Department of Energy ARPA-E to reduce the phenotyping bottleneck that has slowed the improvement of crop varieties (www.terraref.org) (Fig. 1). The field scanner can cover a total of 0.75 hectares of field plots to collect high-resolution, high-frequency, and accurate data on phenotypic traits of interest for diverse sets of germplasm. The scanner system is semi-automated and can scan during day and night hours. It is capable of motion along three axes at velocities of 1.0 m/s on the long axis and 0.33 m/s on the short axis. An array of imaging and sensor systems are integrated with scan motion and data transfer systems, including a thermal infrared camera, hyperspectral imagers, Fraunhofer 3D laser systems, stereo RGB cameras, and NDVI and PRI analog sensors. Additionally, the field phenotyping platform includes a chlorophyll fluorescence imaging system that is a LemnaTec prototype instrument (www.lemnatec.com). A description of the fluorescence imaging system follows in addition to the basic methods for measurements across experimental field plots and a test to determine the impact of variable plant heights. The overall focus is on aspects of field applications of a chlorophyll fluorescence imaging system that differ from analogous systems in controlled environments.

2 Materials

The prototype imaging system is a novel deployment of a chlorophyll fluorescence instrument on a robotic platform in a field setting. The integrated imaging system includes a large panel of red light-emitting diodes (LED). It is operated at night to measure photosystem II activity on dark-adapted plants (Fig. 2). At an operating distance of 80 cm from the plant canopy, the area imaged per measurement is 1.1×0.7 m. The illuminating flash has a dominant wavelength in the range of 620–630 nm and sufficient intensity (up to $7000 \mu\text{mol}/\text{m}^2/\text{s}$ at 70 cm from plant target) to saturate the photosystem of plants. The system is modular and scalable with a controller that interfaces by Ethernet to the camera driver on one end and interfaces with other drivers to synchronize



Fig. 1 The field scanner at the University of Arizona Maricopa Agricultural Center

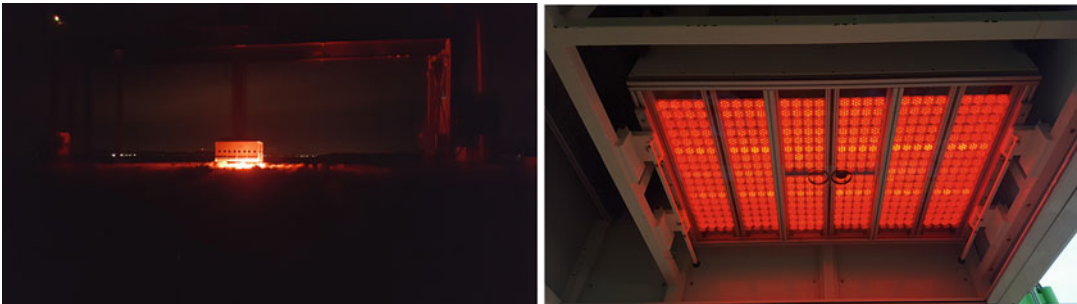


Fig. 2 The chlorophyll fluorescence imaging system includes an LED panel that provides a saturating flash of light over the imaged area

triggering of the flash. The camera is manufactured by Allied Vision (Manta G-235b) and is fit with a bandpass filter mounted on the camera objective in replacement of the UV filter for emitted fluorescence light (690–730 nm, minimum bandwidth 40 nm). The camera is operated at a maximum frame rate of 50 fps, resolution 1936×1216 pixels. The exposure is set at 28 ms with a gain setting of 3000. Data outputs include a series of 101 images captured over 2 s, including the 1 s saturating flash. The images are saved as bin files along with a metadata file that includes the field location and timestamp of the measurement.

3 Methods

3.1 High-Throughput Chlorophyll Fluorescence Imaging of Experimental Field Trials for Dark-Adapted Measurements of F_v/F_m

Procedures for chlorophyll fluorescence imaging in controlled environments are described comprehensively in [10]. The focus here is on the unique aspects of chlorophyll fluorescence imaging in field environments at a scale that accommodates experimental plantings of diversity and mapping panels (~300–900 field plots). While protocols and methods will vary among imaging systems and for different user objectives, several common factors need to be addressed when taking measurements of F_v/F_m on dark-adapted plants in field trials.

1. Plants must be in a fully dark-adapted state prior to the measurement. Restricting scan hours for measurements on dark-adapted plants to nighttime hours 2 h after sunset and before sunrise ensures that plants will be in a dark-adapted state. The intensity of the saturating flash is such that diffuse light disperses to neighboring plants (Fig. 2). Subsequent measurements on plants in the vicinity of a light flash will need to be far enough away to ensure that the measured plants remained fully dark adapted (*see Note 1*).
2. Trigger a saturating flash of light for a duration of 0.8–1 s at an intensity sufficient to saturate the plant photosystems of field-grown plants across the entire imaged area. The illuminating light source needs to provide an irradiance of $>4000 \mu\text{mol}/\text{m}^2/\text{s}$ to capture an accurate measurement of F_m [9, 10]. Results from tests adjusting the illumination intensity in our field scanner system indicate that irradiance of $3500 \mu\text{mol}/\text{m}^2/\text{s}$ is not sufficient, as expected. However, $4200 \mu\text{mol}/\text{m}^2/\text{s}$ is also not quite adequate (*see Note 2*).
3. The saturating flash needs to be synchronized with the camera images and with the programmed robotic motion between measurements. This engineering challenge in the field scanner system is addressed by adjusting the timing of the first image frame to the LED trigger function and programming motion functions that allow for a few seconds delay after the measurement and before movement to a new plot. The requirement for a rapid high-intensity saturating flash creates associated technical challenges in capturing the minimal level of fluorescence (F_0) at exactly the time just after the excitation light is applied [1, 7]. Moreover and perhaps unique to a field setting is that there are sources of background nighttime light that cannot be controlled, for example, moonlight or diffuse light from parking lot lamps (*see Note 3*).
4. A field imaging system needs to be robust against a wide range of environmental conditions. Supplementing the primary components with climate control notably adds to the size and

weight of the complete system. An air conditioner (KTS Hengstberger GmbH) that is able to cool as well as heat is incorporated within the interior chamber above the LED panel that houses the camera and electronics to provide climate control within ambient temperatures in the range from 0 to 50 °C. *See Note 4.*

3.2 Evaluation of the Influence of Variable Plant Heights

Given that the illuminating flash and the fluorescence camera remain at a constant height throughout a data collection scan, it is important to determine the influence of variable plant heights on the data outputs. Differences in plant heights can potentially influence F_v/F_m values as the intensity of the saturating pulse is decreased with increased distance from the light source to the plant target. Additionally, the ability to detect emitted photons may decrease with increased distance from the plant target. To evaluate the influence of variable distances from the plant target to the imaging system, a comparison test can be conducted taking repeat measurements over time on target plots of known height while adjusting the height of the fluorescence camera system. To maintain plants in a dark-adapted state, the measurements need to be separated by spatial distance (>4 m) or by time (~ 30 min or more). Test results completed with an illumination intensity of $4900 \mu\text{mol}/\text{m}^2/\text{s}$ suggest that measurements of mean F_v/F_m on durum wheat at heading stage show a slight decreasing trend as the distance from the plant target to the imaging system increases (simulating a decrease in plant height) in the range from 70 to 140 cm. However, the differences in F_v/F_m are not significant ($p > 0.05$) between standard operational distance (80 cm) compared to distances ranging from 70 to 140 cm (Fig. 3). It is likely that the influence of variable plant heights differs across species with different leaf angles and plant architectures.

4 Notes

1. Preliminary tests on our system indicate that 4 m separation between measurements is sufficient.
2. We find that $4900\text{--}7000 \mu\text{mol}/\text{m}^2/\text{s}$ appears to be sufficient at the standard operating distance from the target.
3. In order to screen out background light, the camera system is synchronized to trigger a single frame before the LED light to establish a baseline intensity in the absence of the red flash. The baseline correction in our system is completed within the camera before images are saved.
4. The instruments need to be protected not only from the external environment but also from internal heating in the electronics and LEDs. Additionally, protection from dust and air

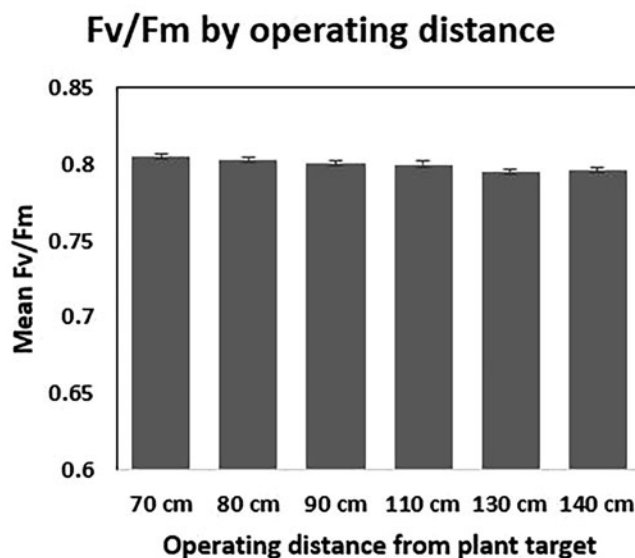


Fig. 3 Graph of mean F_v/F_m as a measure of the quantum efficiency of the PSII system for durum wheat variety Tiburon at heading stage, comparing control plots at a standard operating distance of 80 cm to plots measured at variable operating distances between 70 and 140 cm to simulate variable plant heights. Error bars represent $\pm$ SD

particulates is needed. The air conditioning unit has separated ventilation circuits to allow the clean air inside the chamber to recirculate without mixing with the outside air.

Acknowledgments

The development of methods and tests using the field chlorophyll fluorescence imaging system are the results of a group effort by the larger TERRA-REF project team and especially Tino Dornbusch, Robert Strand, John Huen, Jeffrey Demieville, and Matthew Herritt. This work is part of the TERRA program supported by the Advanced Research Projects Agency-Energy (ARPA-E), US Department of Energy, under Award Number DE-AR0000594. The views and opinions of authors expressed herein do not necessarily state or reflect those of the US government or any agency thereof.

References

1. Murchie EH, Lawson T (2013) Chlorophyll fluorescence analysis: a guide to good practice and understanding some new applications. *J Exp Bot* 64:3983–3998
2. Harbinson J, Prinzenberg AE, Kruijer W et al (2012) High throughput screening with chlorophyll fluorescence imaging and its use in crop improvement. *Curr Opin Biotechnol* 23: 221–226

3. Murchie EH, Kefauver S, Araus JL et al (2018) Measuring the dynamic photosynthetic rate. *Ann Bot* 122:207–220
4. van Bezouw RFHM, Keurentjes JJB, Harbinson J et al (2019) Converging phenomics and genomics to study natural variation in plant photosynthetic efficiency. *Plant J* 97:112–133
5. Oxborough K (2004) Imaging of chlorophyll *a* fluorescence: theoretical and practical aspects of an emerging technique for the monitoring of photosynthetic performance. *J Exp Bot* 55: 1195–1205
6. Baker NR (2008) Chlorophyll fluorescence: a probe of photosynthesis in vivo. *Ann Rev Plant Biol* 59:89–113
7. Henriques FS (2009) Leaf chlorophyll fluorescence: background and fundamentals for plant biologists. *Bot Rev* 75:249–270
8. Björkman O, Demmig B (1987) Photon yield of O₂ evolution and chlorophyll fluorescence characteristics at 77 K among vascular plants of diverse origins. *Planta* 170:489–504
9. Nedbal L, Whitmarsh J (2004) Chlorophyll fluorescence imaging of leaves and fruits. In: Papageorgiou GC, Govindjee R (eds) *Chlorophyll fluorescence: a signature of photosynthesis*. Kluwer Academic Publishers
10. Lawson T, Vialet-Chabrand (2018) Chlorophyll fluorescence imaging. In: Covshoff S (ed) *Photosynthesis methods and protocols*. Humana Press, New York

Part V

Molecular, Metabolomics, Network Analysis, and Quantitative Genetic Analysis of HTP Data



A Method for Rapid and Reliable Molecular Detection of Drought-Response Genes in *Sorghum bicolor* (L.) Moench Roots

Juan B. Fontanet-Manzaneque, David Blasco-Escámez, Damiano Martignago, Andrés Rico-Medina, and Ana I. Caño-Delgado

Abstract

Drought is a major environmental stress that limits growth and productivity in agricultural ecosystems limiting crop yield worldwide. Breeding crops for enhanced drought tolerance is a priority to preserve food security on the increasing world population. Recent work in *Arabidopsis* has shown that vascular brassinosteroid receptor BRL3 (Brassinosteroid insensitive like-3) transcriptionally controls the production of osmoprotectant metabolites that confer drought resistance without penalizing growth, offering new and exciting possibilities for biotechnological improvement of drought-resistant crops. In cereals, understanding transcriptional responses to drought is an essential step for the production of gene-edited drought-resistant cereals. In this chapter, we present a method to analyze the transcriptional responses to drought in *Sorghum bicolor* (L.) Moench, our cereal of choice. Among the genes we tested, we found that drought marker gene *SbDHN1* has a 1000-fold increase only after 1 day of drought, bringing possibilities for the development of molecular sensors for testing drought. Overall, this analysis is useful to set up conditions of high-throughput transcriptomic analysis of drought stressed plants before drought phenotype is observed.

Key words Sorghum, Drought, Root, Drought marker genes, qRT-PCR

1 Introduction

Drought is the most important cause of agricultural losses worldwide, accounting for 29 billion dollars in world agriculture between 2005 and 2015 (FAOSTAT 2018, <http://www.fao.org>). For instance, drought in the United States in 2012 and in Australia in 2006–2008 led to low levels of cereal stock and steep increases in food prices [1]. Consequences of these losses are also affecting developing countries, and drought is a major cause of hunger. In many regions in Asia, rain-fed fields account for 20% of the total rice cultivation area and are vulnerable to the impact of drought [2]. Global population growth, estimated at some 9 billion inhabitants

by 2050, poses a higher pressure on food security [3]. Considering that, engineering drought-resilient crops is one of the most promising solutions to ensure food security in the planet.

Several approaches have been attempted to genetically improve cereals to obtain drought-tolerant varieties. For instance, barley has been genetically modified to increase cytokinin degradation in roots and to reduce stomatal density to improve drought tolerance [4, 5]. Maize has also been genetically modified to delay drought-induced leaf senescence and to reduce the energy consumed in respiration during stresses [6, 7]. Wheat has been genetically engineered to reduce the stomatal density and to obtain higher contents of free proline and soluble sugars, thus maintaining the cellular homeostasis during drought periods [8, 9]. In *Arabidopsis*, our recent work demonstrates that increasing brassinosteroid signaling in the vascular cells confers drought resistance without penalizing growth, by increasing osmoprotectant metabolites in the root [10]. This study prompted us to translate our findings to cereals. We selected sorghum as a cereal of choice to obtain cereal plants with enhanced drought tolerance because it holds several physiological and genetic advantages [11], such as having a diploid and relatively small genome that is amenable to the application of CRISPR-Cas9 for genome editing [12].

Sorghum (*Sorghum bicolor* (L.) Moench) is the fifth most important cereal grass (Poaceae) in terms of production, behind wheat, rice, maize, and barley. Current world production is about 64 million tons of grain from 44.7 million ha. It is grown in 110 countries in all the continents, and the major producers are the United States, Nigeria, Sudan, Mexico, Ethiopia, India, Argentina, China, Niger, and Australia (FAOSTAT 2016, <http://www.fao.org>). Sorghum is a multipurpose crop. It is produced mainly for animal feeding, but it is also produced in developing countries, being consumed as a dietary staple in the form of flat bread and porridges. Nowadays, it is becoming more popular in the brewing industry and to produce bioethanol and bioplastics [13]. From an agronomical perspective, sorghum is one of the better adapted crops to dry environments. This adaptation is mainly based on its biochemical and structural features as C4 photosynthesis, the CO₂-concentrating mechanism in C4 leaves that endows them with higher water use efficiency [14], deep root architecture, and a thick waxy cuticle, maximizing the water uptake from the soil and minimizing the transpiration rate [15]. Some cultivars of sorghum have also a post-anthesis drought resistance, commonly termed stay-green trait, controlled by QTLs [16].

Together with its agronomic and physiological advantages, sorghum offers great genetic advantages for plant breeders. It is a self-pollinating diploid ($2n = 20$) plant with a relatively small genome size (732.2 MB), that is, about 25% of maize genome size (2300 MB) [17]. *Agrobacterium tumefaciens* and particle

bombardment-mediated transformation from immature embryos have been adapted for sorghum [18, 19]. Furthermore, the high degree of synteny among grass genomes led the plant genetic community to view the grass family as a single genetic system [20]; thus, methods presented here could be adapted to other cereals.

Roots are responsible for uptaking water from the soil and transporting it to the aerial plant organs. Roots can rapidly sense the lack of water, and root traits affect the amount of water and nutrient absorption for maintaining crop yield under water stress conditions [21]. The sorghum root architecture is similar to other cereals like maize, being fibrous and composed of seminal roots, which appear at germination, and nodal, crown, or adventitious roots, which emerge later from the shoot [22]. Seminal roots play an important role in initial water and nutrient uptake, whereas nodal roots gain relevance during the later stages of growth and play a fundamental role in plant adaptation to adverse soil conditions [23]. QTLs and genes related with drought resistance have been associated with sorghum root traits, such as nodal root angle [24, 25], but many of these genes, or QTLs, have not been tested toward the genetic engineering of sorghum roots. Therefore, further investigation is required to understand how these genes act in drought-stress conditions in order to evaluate their potential for future editing-guided breeding.

In this chapter, we investigated the transcriptional changes in sorghum roots under laboratory conditions and provide a detailed methodology to set up physiological and transcriptional drought assays in sorghum plants. By using quantitative reverse transcription-polymerase chain reaction (qRT-PCR), a series of genetic markers have been identified that serve to evaluate early drought-stress responses at the molecular level. In brief, we have defined the experimental design and a series of genetic markers to the early detection of drought stress in sorghum. These methods can be scaled to genome-wide expression analyses.

2 Materials

2.1 Sorghum Germination and Growth for Drought Assays

1. *Sorghum bicolor* (L.) Moench M35-1 variety seeds (*see Note 1*).
2. Sterile distilled water.
3. 1 g/L Captan general use fungicide (ethyl mercaptan).
4. Round Petri dishes.
5. Filter paper.
6. Parafilm ®.
7. Rounded tweezers.
8. 65 × 65 × 65 mm square pots.

9. Fertirrigation solution (*see* **Note 2**).
10. Vermiculite substrate.
11. Perlite substrate.

2.2 RNA Purification and Reverse Transcription of Sorghum Root for qRT-PCR

1. Liquid nitrogen.
2. Rounded tweezers.
3. Sterile-RNase free 2 mL tubes.
4. Sterile-RNase free 4 mm Ø glass beads (*see* **Note 3**).
5. Precision balance.
6. Maxwell® 16 LEV Plant RNA kit.
7. NZY First-Strand cDNA Synthesis kit.
8. Centrifuge.
9. Nanodrop™ (ND1000, Thermo Fisher Scientific).
10. TissueLyser™.

2.3 qRT-PCR

1. SYBR Green I.
2. Nuclease-free water.
3. Primers (10 µM).
4. cDNA samples (200 ng/µL).
5. Real-Time PCR 480 LightCycler platform from Roche.
6. qPCR 96-well plates.
7. Optical adhesive sealing sheets.

3 Methods

3.1 Sorghum Germination and Sowing for Drought Assays

1. Place the seeds between two filter papers on liquid medium plates containing Captan solution to avoid fungal growth at the early stages. Keep the seeds in a growth chamber at 28 °C, 60% relative humidity, and photoperiod of 12 h light and 12 h dark.
2. Two days after germination, transfer the seedlings individually to 65 × 65 × 65 mm pots containing 150 g of saturated substrate composed by vermiculite/perlite 6:1 (*see* **Note 4**).
3. As the substrate is inert, fed plants with a fertirrigation solution. Fertirrigation program must be monitored to assure all plants have the same water volume daily.
4. After 5 days of growth in a chamber at 80% relative humidity, photoperiod of 12 h light and 12 h dark, at 28 °C/24 °C, high light regime (150 µmol/m²/s), and fertirrigated, subject plants to drought stress by water-holding, whereas the control plants must remain in the same fertirrigation program.

5. Photograph the plants including an appropriate scale bar to measure phenotyping parameters (root length, plant height, etc.) at each time point of your experimental design (*see* Fig. 1 and Table 1).



Fig. 1 Phenotype of sorghum seedlings without treatment (control) (+) and exposed to drought conditions (withholding water) (–). (Day 1) Seedlings 1 day after applying the drought conditions. (Day 3) Seedlings 3 days after applying the drought conditions. (Day 7) Seedlings 7 days after applying the drought conditions

Table 1
Fertirrigation solution used in the drought experiment. All the nutrients supplied are shown in the first column. The concentrations used of each nutrient are specified in parts per million (PPM). There are also described the minimum, maximum, and optimum concentration values of each nutrient and the soluble form

		Atomic mass	PPM	Minimum	Optimum	Maximum	Soluble form
Nutrients							
Macronutrients	Nitrogen (N)	14.00	231.58	47.00	200.00	400.00	NO ₃ [–] / NH ₄ ⁺
	Phosphorus (P)	31.00	68.38	20.00	50.00	100.00	P ₂ O ₅
	Potassium (K)	39.00	394.94	50.00	400.00	600.00	K ₂ O
	Calcium (Ca)	40.00	172.41	50.00	250.00	500.00	CaO
	Magnesium (Mg)	24.30	29.60	25.00	30.00	150.00	MgO
	Sulfur (S)	32.00	38.98	0.30	400.00	700.00	No soluble form
	Iron (Fe)	55.80	5.40	2.00	5.00	10.00	FeO
Micronutrients	Boron (B)	10.80	0.26	0.25	0.50	5.00	B ₂ O ₃
	Manganese (Mn)	55.00	1.40	0.50	0.80	1.60	MnO
	Zinc (Zn)	65.40	0.28	0.01	0.05	0.50	ZnO
	Molybdenum (Mo)	96.00	0.12	0.01	0.02	0.05	No soluble form
	Copper (Cu)	63.50	0.11	0.01	0.05	0.50	CuO

3.2 RNA Purification of Sorghum Roots for qRT-PCR

1. Separate root tissue samples from the substrate by carefully removing substrate residues by hand. Handle the roots with care to avoid mechanical damage. Do not clean the root sample with water; otherwise, it could invalidate the drought conditions. After cleaning, place the sample rapidly in a tube, and then quickly place it in liquid nitrogen. Samples could be stored at -80°C for further RNA extraction.
2. Add two 4 mm Ø glass beads to each tube containing the samples.
3. Grind the frozen plant material to a fine powder using TissueLyser (TissueLyser II, Qiagen) at 30 Hz for about 4 min (*see Note 5*).
4. Weigh and transfer 60–200 mg of the grinded root tissue into a 2 mL tube.
5. From now the RNA extraction protocol is performed with Maxwell® 16 LEV Plant RNA kit as indicated in the technical manual (Maxwell® 16 LEV Plant RNA kit Technical Manual #TM415). However, some modifications are implemented for sorghum root samples.
6. Add 600 µL of the chilled homogenization solution (Subheading 3.B of the manual), including the L-thioglycerol, to the tube. If plant material is not dissolved in the solution, place tubes on TissueLyser until the sample is homogenized. If foaming occurs, let the sample settle on ice.
7. Transfer 400 µL of the homogenate to a 2 mL tube.
8. Add 200 µL of lysis buffer and vortex vigorously for 15 s.
9. Incubate at room temperature for 10 min.
10. Spin the sample at 20,000 RCF in a centrifuge for 4 min. If the debris is not precipitated, repeat this step.
11. Transfer the supernatant to well #1 of the Maxwell® 16 LEV Cartridge; avoid transferring the precipitate, which might be inadvertently resuspended from the bottom of the tube.
12. Follow the last steps of Maxwell® 16 LEV Plant RNA technical manual, as no more modifications are implemented (Subheading 4 of the manual from **step 10** to the end).
13. Measure RNA concentration in Nanodrop (*see Note 6*).

3.3 Reverse Transcription of Sorghum Root RNA

1. Use 1 µg of RNA for cDNA synthesis using NZY First-Strand cDNA Synthesis kit (NZYTech®) as described in the manufacturer's manual.

Previous analysis by Sudhakar Reddy et al. (2016) selected reference genes which were tested in five different sorghum genotypes, including M35-1 that is used in the present work [26]. Three genes

3.4 Evaluation of Sorghum Reference Genes in Root Tissues Under Drought Stress Conditions for qRT-PCR Data Normalization

3.4.1 Housekeeping Gene Identification

3.4.2 qRT-PCR Analysis for the Selected Housekeeping Genes

resulted the most stable and reliable reference genes in different abiotic stresses and different plant tissues. The genes serine/threonine-protein phosphatase 2A (*SbPP2A*) and cyclophilin/peptidyl-prolyl isomerase (*SbCYP*) were identified as the most stable genes in abiotic stress conditions; and eukaryotic initiation factor 4A-1 (*SbEIF4α*) was the most stable among different tested tissues. However, these conditions did not comprise any drought or osmotic stress. Therefore, a validation experiment needs to be performed to test *SbPP2A*, *SbCYP* and *SbEIF4α* gene expression stability in the experimental conditions used in this work.

The qRT-PCR is performed in 96-well plates using Real-Time PCR 480 LightCycler platform from Roche. The PCR cycling conditions are 95 °C for 10 min, followed by 45 cycles of 95 °C for 10 s and 60 °C for 30 s. Run the qRT-PCR product in a 1% agarose gel to verify primer specificity (*see Note 7*).

3.5 Differential Gene Identification and Experimental Validation to Establish Drought Marker Genes in Sorghum Roots

3.5.1 Gene Identification

Candidate drought marker genes were identified by literature search within various databases and scientific papers. RNA-Seq and cDNA array results from different research articles in sorghum revealed transcriptional activity of more than 28,000 genes differentially expressed under osmotic, dehydration, and ABA-treated conditions [27–29]. From these articles, a series of putative root drought markers were selected, and its transcriptional profile was checked in Morokoshi Sorghum transcriptional database [30]. Finally, genes which were transcriptionally expressed in seedling roots and differentially expressed between these and Root_PEG and Root_ABA datasets were selected to perform an experimental validation. The selected genes were (Fig. 2):

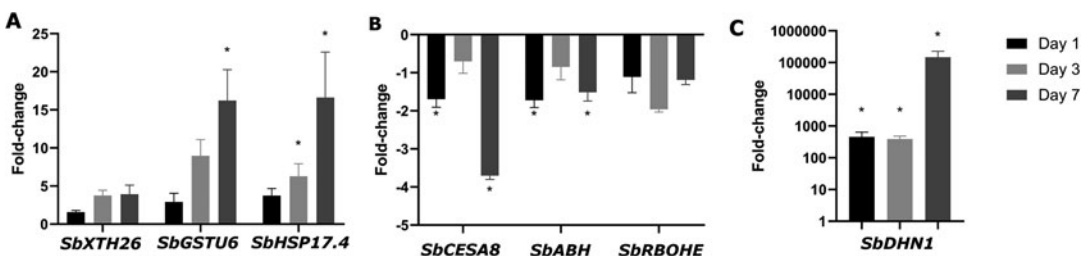


Fig. 2 qRT-PCR analysis in sorghum roots for relative expression of the putative markers under drought stress conditions at different time points (1, 3, and 7 days after drought) where the *SbCYP* gene was used as an internal control to normalize the expression level. (a) Relative gene expression of upregulated markers. (b) Relative gene expression of downregulated markers. (c) Relative gene expression of highly sensitive marker *Sorghum bicolor* dehydrin (*SbDHN1*). The *SbDHN1* fold change is reaching values of 1000 to 600,000 because of the low expression levels of this gene in basal conditions, with an absolute value of 1.03×10^{-5} . Bars and asterisks denote standard errors and significant differences against unstressed plants using a Student's t-test of three independent biological replicates (P value < 0.05), respectively

- *Sorghum bicolor* dehydrin 1 (*SbDHN1*): *Sb09g018420*
- *Sorghum bicolor* xyloglucan endotransglucosylase/hydrolase (*SbXTH26*): *Sb04g010980*
- *Sorghum bicolor* heat shock protein 17.4 KDa (*SbHSP17.4*): *Sb01g039990*
- *Sorghum bicolor* glutathione S-transferase U6 (*SbGSTU6*): *Sb03g031780*
- *Sorghum bicolor* cellulose synthase A catalytic subunit 8 (*SbCESA8*): *Sb02g007810*
- *Sorghum bicolor* alpha/beta hydrolase (*SbABH*): *Sb01g045300*
- *Sorghum bicolor* respiratory burst oxidase E (*SbRBOHE*): *Sb02g025660*

3.5.2 Primer Design

Primers were designed using Primer-BLAST (NCBI). Optimum primer parameters for qPCR were PCR product size from 70 to 200 bp with a primer melting temperature from 57 to 63 °C, being the optimum at 60 °C and with a maximum melting temperature difference of 1 °C. Exon junction span is prioritized when possible (*see Note 8*). Primer pair specificity to the intended PCR template should also be checked (*see Note 9*). Primer pair sequences are specified in Table 2.

3.5.3 qRT-PCR Primer Efficiency Standard Curve Analysis

Primers pairs should be tested using a standard curve. To perform a qPCR standard curve, set up a qPCR reaction to amplify different amounts of the same cDNA sample; efficient primers will result in a proportional dose-response curve. Test four different concentrations with a dilution factor of 1:5, doing sequential dilutions. To detect contaminations in the reaction and to discriminate background amplification, a negative control must be included, using nuclease-free water.

3.5.4 qRT-PCR Analysis for the Selected Putative Drought Marker Genes in Sorghum Roots

Analyze the expression of the selected putative drought marker genes under drought stress treatment by qRT-PCR. Use the *SbCYP* gene as an internal control to normalize the sample amounts (*see Note 10*). qRT-PCR is performed in 96-well plates using Real-Time PCR 480 LightCycler platform (Roche). PCR cycling conditions are 95 °C for 10 min, followed by 45 cycles of 95 °C for 10 s and 60 °C for 30 s (*see Note 11*). qRT-PCR product was run in an agarose gel to verify primer specificity.

3.6 Mathematical Method for Relative Quantification of qRT-PCR Data ($2^{-\Delta\Delta ct}$)

The relative quantification is calculated using the $2^{-\Delta\Delta ct}$ method [31], in which *ct* indicates cycle threshold. $\Delta\Delta ct = (ct_{\text{drought marker (stressed plants)}} - ct_{\text{SbCYP (stressed plants)}}) - (ct_{\text{drought marker (control plants)}} - ct_{\text{SbCYP (control plants)}})$.

Table 2**Details of reference and drought marker genes, nomenclature, primer sequences, product size, and melting temperature**

Gene annotation	Oligo sequence	Amplicon length	T _m (°C)
Sb04g007570 (<i>SbPP2A</i>)	Fw-AACCCGCAAAACCCCAGACTA Rv-TACAGGTCGGGCTCATGGAAC	138	59
Sb04g019590 (<i>SbCYP</i>)	Fw-GTATCTGTGCTCGCCGTCTCT Rv-TTCACCCAACTCCTCAACCCC	108	59
Sb04g003390 (<i>SbEIF4α</i>)	Fw-CAACTTTGTCACCCGCGATGA Rv-TCCAGAAACCTTAGCAGCCCA	144	58.3
Sb09g018420 (<i>SbDhm1</i>)	Fw-CCACAAGGACAACCAGCAC Rv-TTCACACGCCAGAGAGAGC	191	57
Sb04g010980 (<i>SbXTH26</i>)	Fw-TGACACCAGGAAGGACAGTG Rv-TGGAACTAAACACCCCCAAA	114	55.5
Sb01g039990 (<i>SbHSP17.4</i>)	Fw-CGACATCAAGAACGTCCAGA Rv-CGAAAGCACGTCCCTTTTAC	100	54.4
Sb03g031780 (<i>SbGSTU6</i>)	Fw-CATCGACGACAAGTTTATCCTG Rv-CTCCAACGTCTCCATCACG	109	54.4
Sb02g007810 (<i>SbCesA8</i>)	Fw-CGGCCAGATGGTTGATGACA Rv-GAGGTTGGGATCTGCGAAGG	118	57.8
Sb01g045300 (<i>SbABH</i>)	Fw-CACCGATACCTCAGGACGTG Rv-GACCTTGTTGAAGCGAGCAG	170	57
Sb02g025660 (<i>SbRbohE</i>)	Fw-AGTACAGGAGCTGATCGTGC Rv-ATGTAGCCAAGGTTCTCGGG	120	57

4 Notes

1. Accession number of M35-1 variety seeds: “PI656047.”
2. The fertirrigation program is specified in Table 1.
3. For the removal of RNase, the glass beads are treated with 0.1 N NaOH overnight. Then the beads are rinsed with distilled water and kept at 200 °C for 4 h.
4. An inert substrate composed by vermiculite and perlite 6:1 was used to facilitate the cleaning of the root samples for RNA extraction as vermiculite is a fine substrate and it is easily removed by hand.
5. For grinding root samples, it is preferable to use 2 mL Eppendorf, and glass beads should have a diameter bigger than 4 mm, due to the thickness of cereal roots. In addition, when grinding the root plant material in the TissueLyser, it is necessary to do it in 2 min intervals. Meanwhile, keep the samples in liquid nitrogen after each interval to prevent RNA degradation.

6. The RNA concentration is expected to be in the range from 200 to 500 ng/ μ L and the ratio of absorbance at 260 nm and 280 nm (260/280 ratio) from 2 to 2.2.
7. Primer sequences used to test the housekeeping genes by qRT-PCR were taken from Sudhakar Reddy et al. (2016) article [24], specified in Table 2.
8. It is advisable that primers span an exon-exon junction to be specific for the amplification of cDNA and therefore not amplifying genomic DNA residues.
9. When checking the primer specificity to your PCR template, you can narrow down your search to *Sorghum bicolor* RefSeq mRNA database (taxid: 4558) present in Primer-BLAST tool.
10. *SbCYP* is used as an internal control because it was the most stable one in our experimental conditions.
11. The *SbDHN1* is the most recommendable drought marker gene due to its high sensitivity, being strongly upregulated even before the plants showed any phenotypical change. For early drought responses, the drought markers *SbABH* and *SbCESA8* are recommendable as well, due to their significantly twofold decrease at 1 day after applying the stress. For later drought responses, the use of the marker *SbHSP17.4* is proposed because of its 5- and 15-fold increase at 3 and 7 days after applying the stress, respectively.

References

1. Iizumi T, Ramankutty N (2015) How do weather and climate influence cropping area and intensity? *Glob Food Sec* 4:46–50
2. Nakashima K, Suenaga K (2017) Toward the genetic improvement of drought tolerance in crops. *JARQ* 51(1):1–10
3. The World Bank (2007) World Development Report 2008. <https://doi.org/10.1596/978-0-8213-6807-7>
4. Ramireddy E, Hosseini SA, Eggert K et al (2018) Root engineering in barley: increasing cytokinin degradation produces a larger root system, mineral enrichment in the shoot and improved drought tolerance. *Plant Physiol* 177(3):1078–1095
5. Hughes J, Hepworth C, Dutton C et al (2017) Reducing stomatal density in barley improves drought tolerance without impacting on yield. *Plant Physiol* 174(2):776–787
6. Bedada LT, Seth MS, Runo SM et al (2016) Drought tolerant tropical maize (*Zea mays* L.) developed through genetic transformation with isopentenyltransferase gene. *Afr J Biotech* 15(43):2447–2464
7. Njuguna E, Coussens G, Aesaert S et al (2017) Modulation of energy homeostasis in maize and Arabidopsis to develop lines tolerant to drought, genotoxic and oxidative stresses. In *Afrika Focus* 30(2). <https://doi.org/10.21825/af.v30i2.8080>
8. Bi H, Shi J, Kovalchuk N et al (2018) Overexpression of the TaSHN1 transcription factor in bread wheat leads to leaf surface modifications, improved drought tolerance, and no yield penalty under controlled growth conditions. *Plant Cell Environ* 41(11):2549–2566
9. Gao H, Wang Y, Xu P et al (2018) Overexpression of a WRKY transcription factor TaWRKY2 enhances drought stress tolerance in transgenic wheat. *Front Plant Sci* 9. <https://doi.org/10.3389/fpls.2018.00997>
10. Fabregas N, Lozano-Elena F, Blasco-Escamez D et al (2018) Overexpression of the vascular brassinosteroid receptor BRL3 confers drought resistance without penalizing plant growth. *Nat Commun* 9(1):4680
11. Blasco-Escamez D, Lozano-Elena F, Fabregas N et al (2017) The primary root of *Sorghum*

- bicolor* (L. Moench) as a model system to study brassinosteroid signaling in crops. In: Brassinosteroids. Humana Press, New York
12. Li A, Jia S, Yobi A et al (2018) Editing of an alpha-kafirin gene family increases digestibility and protein quality in sorghum. *Plant Physiol* 177(4):1425–1438
 13. Taylor JR, Schober TJ, Bean SR (2006) Novel food and non-food uses for sorghum and millets. *J Cereal Sci* 44(3):252–271
 14. Ghannoum O (2008) C4 photosynthesis and water stress. *Ann Bot* 103(4):635–644
 15. Rostamza M, Richards RA, Watt M (2013) Response of millet and sorghum to a varying water supply around the primary and nodal roots. *Ann Bot* 112(2):439–446
 16. Buchanan CD, Lim S, Salzman RA et al (2005) *Sorghum bicolor*'s transcriptome response to dehydration, high salinity and ABA. *Plant Molec Biol* 58(5):699–720
 17. Paterson AH, Bowers JE, Bruggmann R et al (2009) The *Sorghum bicolor* genome and the diversification of grasses. *Nature* 457(7229):551
 18. Howe A, Sato S, Dweikat I et al (2006) Rapid and reproducible *Agrobacterium*-mediated transformation of sorghum. *Plant Cell Rep* 25(8):784–791
 19. Belide S, Vanhercke T, Petrie JR et al (2017) Robust genetic transformation of sorghum (*Sorghum bicolor* L.) using differentiating embryogenic callus induced from immature embryos. *Plant Methods* 13(1):109
 20. Hamblin MT, Casa AM, Sun H et al (2006) Challenges of detecting directional selection after a bottleneck: lessons from *Sorghum bicolor*. *Genetics* 173(2):953–964
 21. Wasaya A, Zhang X, Fang Q et al (2018) Root phenotyping for drought tolerance: a review. *Agronomy* 8(11):241
 22. Singh V, van Oosterom EJ, Jordan DR et al (2010) Morphological and architectural development of root systems in sorghum and maize. *Plant Soil* 333(1–2):287–299
 23. Smith S, De Smet I (2012) Root system architecture: insights from Arabidopsis and cereal crops. *Phil Trans R Soc B* 367:1441–1452
 24. Mace ES, Singh V, Van Oosterom EJ et al (2012) QTL for nodal root angle in sorghum (*Sorghum bicolor* L. Moench) co-locate with QTL for traits associated with drought adaptation. *Theoret Appl Genet* 124(1):97–109
 25. Harris-Shultz KR, Hayes CM, Knoll JE (2019) Mapping QTLs and identification of genes associated with drought resistance in sorghum. In: *Sorghum*. Humana Press, New York
 26. Sudhakar RP, Srinivas RD, Sivasakthi K et al (2016) Evaluation of sorghum [*Sorghum bicolor* (L.)] reference genes in various tissues and under abiotic stress conditions for quantitative real-time PCR data normalization. *Front Plant Sci* 7:529
 27. Dugas DV, Monaco MK, Olson A et al (2011) Functional annotation of the transcriptome of *Sorghum bicolor* in response to osmotic stress and abscisic acid. *BMC Genomics* 12(1):514
 28. Buchanan CD, Lim S, Salzman RA et al (2005) *Sorghum bicolor*'s transcriptome response to dehydration, high salinity and ABA. *Plant Molec Biol* 58(5):699–720
 29. Johnson SM, Lim FL, Finkler A et al (2014) Transcriptomic analysis of *Sorghum bicolor* responding to combined heat and drought stress. *BMC Genomics* 15(1):456
 30. Makita Y, Shimada S, Kawashima M et al (2014) MOROKOSHI: transcriptome database in *Sorghum bicolor*. *Plant Cell Physiol* 56(1):e6–e6
 31. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2[−] $\Delta\Delta$ CT method. *Methods* 25(4):402–408



High-Throughput Profiling of Metabolic Phenotypes Using High-Resolution GC-MS

Nishikant Wase, Nathan Abshire, and Toshihiro Obata

Abstract

Metabolite profiling provides insights into the metabolic signatures, which themselves are considered as phenotypes closely related to the agronomic and phenotypic traits such as yield, nutritional values, stress resistance, and nutrient use efficiency. GC-MS is a sensitive and high-throughput analytical platform and has been proved to be a vital tool for the analysis of primary metabolism to provide an overview of cellular and organismal metabolic status. The potential of GC-MS metabolite profiling as a tool for detecting metabolic changes in plants grown in a high-throughput plant phenotyping platform was explored. In this chapter, we describe an integrated workflow of semi-targeted GC-high-resolution (HR)-time-of-flight (TOF)-MS metabolomics with both the analytical and computational steps, focusing mainly on the sample preparation, GC-HR-TOF-MS analysis part, and data analysis for plant phenotyping efforts.

Key words Metabolomics, GC-HR-TOF-MS, Plant phenotyping, Phenomics

1 Introduction

In recent years, plant phenotyping has received increased interest to bridge the gap between genotype-to-phenotype knowledge. At present, there is an urgent need for expanding the phenotypic capabilities to measure more complex phenotypic traits. As aptly quipped by Tuberosa, “Phenotyping is king and heritability is queen” [1]. The classical plant phenotyping tools generally rely on manual measurement of selected traits from limited sampling of plants, have limited throughput, and hence are the main hurdle in the comprehensive analysis of phenotypic traits within a single plant species and different cultivars. This is sometimes also referred to as a phenotyping bottleneck [2]. To functionally describe the metabolism of plants, metabolomics has turned out to be a powerful and indispensable tool.

Metabolism can be defined as the large array of chemical reactions taking place inside the cell. Enzymes present in the cell guide

these reactions, and the resultant products are called metabolites. Metabolites are the products of essential metabolic processes such as respiration, photosynthesis, and biosynthesis of building blocks for macromolecules. Metabolites such as proteinogenic amino acids, nucleotides, lipids, and carbohydrates can be categorized as primary metabolites which have close relationships with crop performances due to their essential roles in growth, development, and life cycle. On the other hand, secondary metabolites play important roles in the interaction with the environment. Thus, a plant's metabolite profile is a component that can bridge the genotype-phenotype gap and facilitate the selection of superior traits in plant breeding programs [3]. Metabolomics is an excellent tool to characterize metabolic phenotype rigorously and link the traits to associated gene expression changes and changes in the phenotype and improve our understanding of the genotype-phenotypic association of desired traits [4]. Hence, metabolomics can be largely viewed as a physiological analysis enabling us to come to a biological conclusion of plant productivity and interaction with abiotic and environmental stressors [5]. An excellent example is a recent study by Steinfath et al. using metabolite profiling to predict the phenotypes of agriculture importance [6]. In this study, 20 potato cultivars were analyzed to identify biomarkers responsible for the susceptibility of potato to mechanical damage and browning of potato during frying due to Millard's reaction. Authors found that serine, valine, tyrosine, threonine, and glutamine are the biomarkers for mechanical bruising, while glucose and fructose are identified as biomarkers for browning of potato during frying [7]. Thus, metabolite profiling can also be used to predict crop product quality and identify biomarker metabolites which can be used for diagnostics and marker-assisted breeding [8].

Metabolic profiling studies have been performed using a range of analytical platforms, including gas chromatography-mass spectrometry (GC-MS) and liquid chromatography (LC)-MS. There are other variants such as capillary electrophoresis (CE)-MS, nuclear magnetic resonance (NMR) spectroscopy, or direct infusion MS. Because of the diversity of the physical and chemical properties of the metabolites, no single analytical platform can profile all metabolites in a sample [9]. Therefore, the platform must be chosen depending on the classes of metabolites to be analyzed. The most widely used methods are GC- or LC-MS and NMR, and each offers different advantages and disadvantages [10]. The main strength of GC-MS, which makes it suitable for high-throughput metabolite profiling, is its reproducibility and wide dynamic range. These allow us to compare the data from different analytical batches and to minimize the possibility of saturation and misdetection. Additionally, the GC retention time and MS fragmentation pattern information are reproducible and can be shared between instruments even when they are using slightly

different settings. This makes the semi-targeted metabolite profiling possible by using publicly available data libraries typically containing the information of thousands of metabolites belonging to a wide variety of chemical classes. Especially the GC retention time information makes the peak annotation reliable. Electron ionization (EI) fragmentation method fragments molecules very reproducibly according to their structure which helps in identification and overrides the need for tandem mass spectrometry. For these reasons, GC-MS has been widely used for profiling of the primary metabolites, as it can quantify compounds of molecular weight smaller than 600 a.u. Using a high-resolution (HR)-time-of-flight (TOF)-MS is beneficial over conventional GC-quadrupole (Q)-MS whose resolution is limited to the nominal mass or up to one decimal place. With the GC-HR-TOF-MS, the data can be acquired at a high full-scan rate of up to 50 spectra/sec, high mass accuracy and resolution with four decimal places, the increased limit of detection, and enhanced measurement precision [11]. GC-HR-TOF-MS also has an excellent ability to separate ion signals at different exact masses but equal nominal masses, which greatly contribute to avoiding incorrect annotation of peaks and eliminating the effects of co-eluting metabolites on peak quantification. For example, a characteristic fragment of citric acid (nominal mass of the monoisotopic fragment is 273.0 m/z) is detected as a single fragment by GC-Q-MS in a plant sample. However, two fragments with m/z 273.0843 and 273.1207 were detected using GC-HR-TOF-MS (Fig. 1). The fragment with m/z 273.1207 is not derived from citrate but most likely from a metabolite co-eluted with it. These two fragments cannot be separated by GC-Q-MS, leading to an error in the quantification of citrate. Additionally, the GC-HR-TOF-MS system has the ability to detect and identify trace-level metabolites within a complex matrix. A GC-Q-MS in selective ion monitoring (SIM) mode or GC triple quadrupole mass spectrometer operating in multiple reaction monitoring (MRM) mode offers better selectivity and dynamic range, but these techniques are optimal for targeted analysis. Thus, using the GC-HR-TOF-MS system for phenotyping analysis has clear advantages over the other GC-MS systems.

Metabolite profiling is a multistep procedure, and an optimal workflow is required for various steps, including preparatory, analytical, and computational steps [12, 13]. The preparatory steps involve paying due diligence to the experimental design, considering both biological and analytical constraints, sample collection, sample handling, metabolic quenching, and storage of the material. As a rule of thumb, special care should be given to the sample collection and handling routines in any metabolite profiling analysis to minimize perturbations of the sample's physiological state. For example, harvest time in the day, shading, and wounding can significantly affect levels of many metabolites. Moreover,

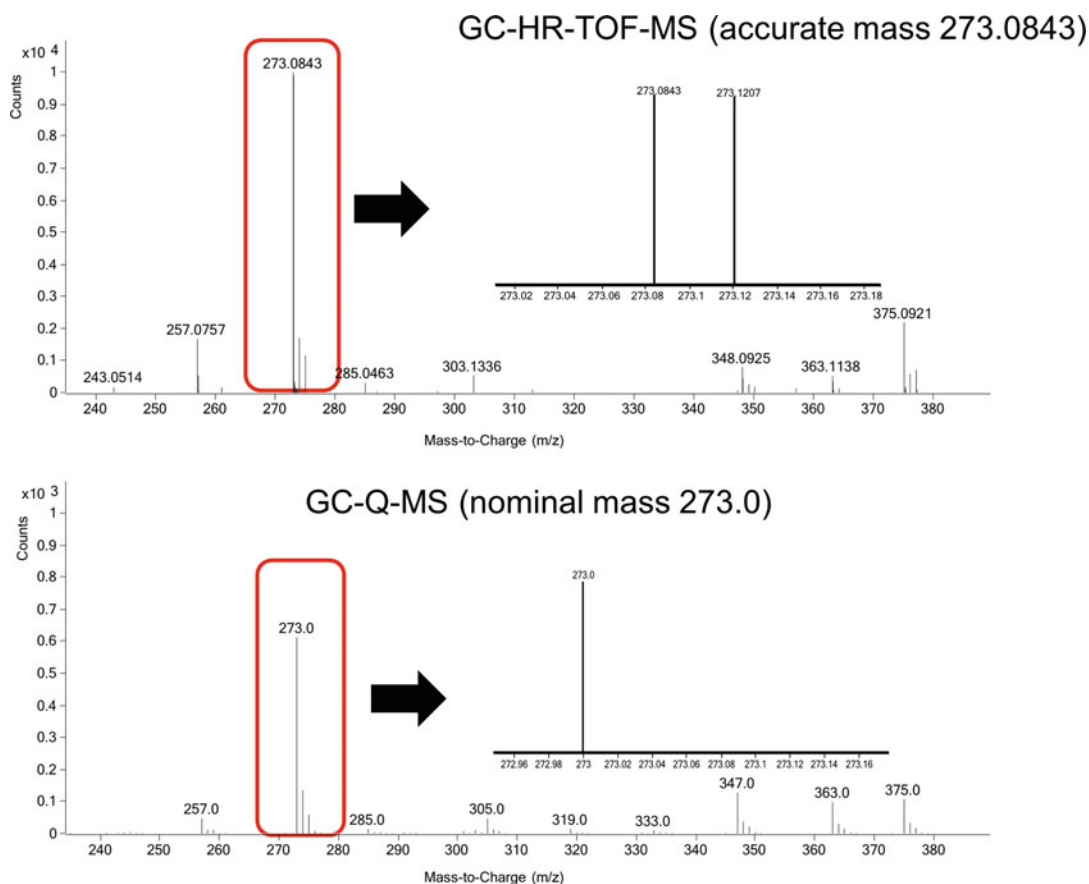


Fig. 1 Comparison between high- and low-resolution GC-MS analyses. The upper and lower panels show parts of mass spectra of citric acid in a biological sample acquired on high- and low-resolution GC-MS analysis, respectively. The ion with m/z 273.0843 is detected by HR-MS separately from the one with m/z 273.1207, which is not derived from citric acid, while these are recognized as a single ion by a conventional MS

optimization of this step is very critical considering the fact that metabolism is a very dynamic cellular process. Quenching is generally carried out by flash freezing of the sample in liquid nitrogen. Samples should be stored at -80°C before analysis and transported on dry ice. An ideal metabolite extraction should satisfy three criteria [14, 15]: (i) method should completely extract all intracellular metabolite pools, (ii) metabolite conversion should be prevented or minimized during the extraction or subsequent steps and this can be achieved by effective inactivation of enzymes, and (iii) metabolite degradation should be minimized. Although there is no single universal extraction method for the entire metabolome, extraction of water-soluble metabolites can be achieved using organic solvents, high/low temperature, mechanical cracking, extreme pH, or combinations of these [16–21]. In the 1990s, biphasic liquid extraction with chloroform/methanol was

introduced [14] and has been the mainstay of the metabolite extraction procedures and employed in the protocol introduced here too.

Once the data is acquired by the GC-HR-TOF-MS, the computational analysis of data processing can be divided into three main sections:

- (a) Peak detection by spectral deconvolution and identification of the peaks of metabolites using MassHunter Unknowns Analysis
- (b) Quantification of the identified metabolite peaks by MassHunter Quantitative Analysis
- (c) Statistical analysis and biological interpretation

The entire workflow is shown in Fig. 2. Once the GC-HR-TOF-MS data for a batch is acquired, the first step in our workflow is the identification of the compound peaks using Unknowns Analysis. This is primarily done using MassHunter Unknowns Analysis software. A mix of the samples (pooled reference sample) is used for the identification of the metabolite peaks assuming that the reference sample represents all the samples and should contain all the metabolites present in samples under consideration. The data is deconvoluted, and resulting fragmental peaks are aligned to generate “components” which are sets of fragmental ion peaks supposed to be derived from the single metabolites. The components are then compared with a standard spectral library, i.e., Fiehn GC-MS mass spectral library [22], which contains GC retention index (RI) information for each metabolite. The spectral search and RI matching result in a peak list with library hits (tentatively identified) and non-hits (not identified) peaks. The library hits are then exported out as a reference library specific for the analytical batch after a rigorous curation process (*see* Subheading 3.5, **Notes 8 to 15**). All the samples, including the blanks and reference samples, are then loaded into MassHunter Quantitative Analysis software as a batch. Batch-specific quantitation method is created based on the reference library generated by the MassHunter Unknowns Analysis software. The quantitation software matches the retention time and mass spectra of the target peaks present in the reference library to those of peaks present in each sample to identify metabolite peaks. Ion intensity of the “quantification ion,” which represents the abundance of the metabolite, is determined as the peak height for each detected metabolite in each sample. Thus, starting from raw GC-TOF peaks, the user can confidently identify metabolite peaks in the complex mixture and quantify its peak height. These peak height values can further be normalized and used for statistical analysis to draw biological insight related to the phenotypic traits as discussed in Subheading 3.7.

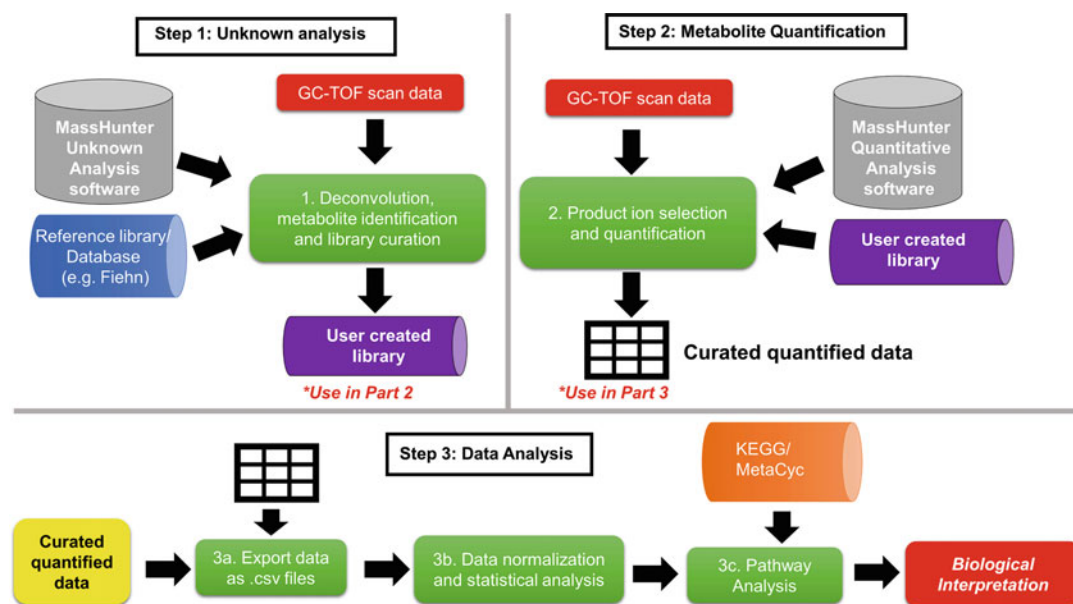


Fig. 2 Overview of the data analysis workflow. This workflow is broadly divided into three main steps. **Step 1** is performed using MassHunter Unknowns Analysis software to identify the metabolite peaks to be analyzed in the batch. The identified compounds are then exported as a reference library and used in the quantification workflow (**step 2**). The quantification part starts with creating a quantification method in MassHunter Quantitative Analysis software using the user-created reference library in **step 1**. Metabolite peaks are identified and quantified in each sample. Generated peak height data for individual metabolites are filtered and normalized to represent relative metabolite levels and used for statistical analysis to draw biological interpretation

Taking all the considerations, we present a protocol for data analysis that can be applied to any biological sample (with some adaptation). The protocol integrates the experimental and computational parts. The experimental part deals with polar metabolite extraction, derivatization, and GC-HR-TOF-MS data acquisition and then metabolite peak identification and quantification using MassHunter software suit. Further data processing practice including data normalization, filtering, and statistical analysis will also be described. Our research group routinely uses this protocol to acquire profiles of plants, animal and human tissues, cell cultures, and other biological samples.

2 Materials

2.1 Consumables

1. Two mL microcentrifuge tubes (Eppendorf; Cat. # 022363352 (*see Note 1*)).
2. Pipette tips.
3. One and a half mL autosampler glass vials with PTFE crimp cap (Agilent Cat # 5181-3376 or similar) with 200 μ L glass insert with plastic spring (Cat # 5181-1270).

4. Argon gas (purity 99.99%; Cat # AR UHP 1L, MATHESON, 166 Keystone Drive, Montgomeryville, PA).
5. Liquid nitrogen. Liquid nitrogen is a cryogenic liquid. It is liquefied under high-pressure and low-temperature conditions and can expand to a very large volume of gas. Liquid nitrogen and its vapor can rapidly freeze skin tissues and eye fluids, resulting in a cold burn, frostbite, and permanent damage even by brief exposure. It also causes asphyxiation as nitrogen can replace oxygen in the air. Always use only approved containers that can withstand low temperatures. Always wear eye/-face protection, skin protection, thermally insulated leather gloves, long sleeve shirts, and safety shoes.

2.2 Reagents

1. HPLC-grade water.
2. HPLC-grade methanol (Sigma; Cat # 34885). Methanol is toxic and highly flammable and should be handled in a chemical hood.
3. HPLC-grade chloroform (EMD Millipore; Cat # CX1050). Chloroform is toxic and highly flammable and should be handled in a chemical hood.
4. Ribitol/adonitol (Sigma; Cat # A5502) (as internal standard).
5. Pyridine anhydrous (EMD Millipore; Cat # PX2012). Pyridine is harmful and highly flammable and should be handled in a chemical hood.
6. Methoxyamine hydrochloride (Sigma; Cat # 89803). Methoxyamine HCl is corrosive and harmful and should be handled in a chemical hood.
7. MSTFA (N-methyl-N-(trimethylsilyl)trifluoroacetamide; 98 + %) (CovaChem; Cat # 12104-10x1). Methoxyamine HCl is irritant and flammable and should be handled in a chemical hood.
8. Fatty acid methyl esters (FAMES) mixture containing C:8 to C:30 FAMES. Weigh 10 mg each in 1.5 mL Eppendorf tubes, and add 1 mL of chloroform. Combine all of the stocks in a clean glass bottle, and make up the volume 25 mL with chloroform. Aliquot into glass vials and store at -20°C . Chemicals are irritant, wear appropriate personal protection, and perform operations in a chemical hood. [Methyl caprylate (C8); methyl pelargonate (C9); methyl caprate (C10); methyl laurate (C12); methyl myristate (C14); methyl palmitate (C16); methyl stearate (C18); methyl eicosanoate (C20); methyl docosanoate (C22); lignoceric acid methyl ester (C24); methyl hexacosanoate (C26); methyl octacosanoate (C28); triacontanoic acid methyl ester (C30).]

2.3 Equipment

1. Microcentrifuge.
2. Mortar and pestle (for the homogenization of a small amount of plant tissue such as leaf/stem/root).
3. Stainless steel Dewar flask for liquid nitrogen.
4. Weighing balance.
5. Centrifugal vacuum concentrator with cold trap (e.g., Centri-Vap, Labconco Corp. Kansas City, MO).
6. Thermomixer (i.e., MultiTherm, Benchmark Scientific).
7. TissueLyser II (Qiagen, Germantown, MD).
8. Vortex mixer.

2.4 GC-HR-TOF-MS System

1. 7890B GC system (Agilent Technologies, Santa Clara, CA).
2. 7200 series GC-QTOF system (Agilent Technologies, Santa Clara, CA).
3. 7893 autosampler (Agilent Technologies, Santa Clara, CA).
4. GC capillary column: Agilent J&W HP-5MS, 30 m $\times$ 0.25 mm $\times$ ID 0.25 μ m.

2.5 Software

1. MassHunter GC/MS Acquisition (Agilent Technologies, Santa Clara, CA).
2. MassHunter Workstation VB.08.00 (Agilent Technologies, Santa Clara, CA).
3. Qualitative Navigator module VB.08.00 (Agilent Technologies, Santa Clara, CA).
4. Quantitative Analysis module VB.08.00 (Agilent Technologies, Santa Clara, CA).
5. Quantitative Unknowns Analysis module VB.09.00 (Agilent Technologies, Santa Clara, CA).
6. Agilent Fiehn GC/MS Metabolomics RTL Library (2013) [22].
7. Microsoft Excel 2015 (Microsoft, Redmond, WA, USA).

3 Methods**3.1 Sample Preparation and Metabolite Extraction**

1. The tissue should be pulverized to a fine powder under liquid nitrogen temperature using TissueLyser or mortar and pestle. Take 50–100 mg (*see Note 2*) of pulverized material in a 2 mL Eppendorf microfuge tube (round bottom shaped). Weigh the plant material with precision. The precision on sample weight is essential as it would be later used for normalization. Perform this operation under liquid nitrogen temperature.

2. Add 730 μL of methanol premixed with 20 mg/mL ribitol in water at 700:30 ratio (*see* **Note 3** for optimization of internal standard concentration), and vortex immediately and vigorously. Keep the samples on ice.
3. Transfer the tubes to a thermomixer set at 70 °C, and shake the tubes for 15 min at 950 rpm. After 1–2 min, open the Eppendorf tubes for a moment to release the extra pressure. All enzymatic activities should be stopped here, and the following procedures can be performed at RT.
4. Centrifuge 10 min at $17,000 \times g$ and transfer the supernatant (without disturbing the pellet) to a new tube. Add 325 μL chloroform and 750 μL water, and vortex the tubes vigorously for 30 sec.
5. Centrifuge the tubes for 15 min at $1500 \times g$. Take an aliquot of 50 μL from the upper polar phase into a fresh 2 mL Eppendorf tube (*see* **Note 4** for optimization for the volume of metabolite extract). Dry the samples with a centrifugal vacuum concentrator. The rest of the supernatant can be divided into 2–4 aliquots and kept as backup samples after drying. The tubes are filled with argon gas immediately after vacuum drying and are tightly closed to prevent the oxidation of metabolites. The samples can be stored at either RT or –80 °C for several months.

3.2 Derivatization

GC-MS-based metabolite profiling requires the derivatization of metabolites to make them volatile and thermally stable [23]. The derivatization procedure is a two-step procedure involving (a) the methoximation of the metabolites containing aldehyde and ketone group to inhibit ring formation and (b) the silylation of all metabolites from the previous step to form TMS derivatives [24]. Although methoximation and silylation of the metabolites are effective procedures, they are also marred with errors in GC-MS data variance in thermal stability of the metabolites, derivatization efficiency, and generation of multiple peaks [25–27].

1. Vacuum dry stored samples for 30 minutes to remove any residual humidity (*see* **Note 5**).
2. Add 40 μL methoxyamine hydrochloride solution (20 mg/mL in pyridine). When this chemical is not well dissolved in pyridine, warm the solution by hand. Follow the procedures below with one additional empty tube to be analyzed as “blank.”
3. Incubate the tubes at 37 °C for 2 h with shaking at 950 rpm.
4. Add 70 μL of MSTFA mix (1 mL of MSTFA + 20 μL of FAMES mix), and incubate at 37 °C for 30 min with shaking at 950 rpm.
5. Create a pooled reference sample by aliquoting 10 μL of derivatized material from each sample into a separate tube.
6. Transfer the derivatized metabolites to autosampler vials with a glass insert, and seal the vials with a crimper.

3.3 GC-MS Data Acquisition

Set the inlet temperature to 230 °C and the auxiliary transfer line temperature to 280 °C. The GC oven temperature gradient program is as follows: set the initial oven temperature to 80 °C and hold for 2 min, and then ramp the temperature up to 330 °C at a rate of 15 °C/min. Finally, hold the temperature for 6 min. The total runtime is 24.6 min. The helium carrier gas flow is set to 1 mL/min. Operate the mass spectrometer over a scan range of 60–600 m/z at -70 eV. Provide a solvent delay of 5 min so that the injection peak data is skipped. Inject 1 μ L metabolite mixture to acquire its metabolic profile (*see* **Note 6**). Include TOF calibration at the beginning of the sequence and after every 20 to 30 samples to gain accurate m/z values (add empty sample with Type column “Keyword” and Keyword column “MassCal” in the sequence).

3.4 Generation of Experiment-Specific Retention Index (RI) Calibration File

Calculations of relative retention were adapted for temperature programmed GC separations, and these values are known as RI. This allows us the comparison between the retention times in the library and in own analyses even if those are based on different analytical platforms. RI calculation requires a mixture of standard analytes (in the current protocol, FAMES mixture with C8 to C30) that covers the elution range of the metabolites of interest. Each FAME species is assigned an RI value based on the number of carbon atoms multiplied by 100, e.g., methyl caprylate (C8), methyl pelargonate (C9), and methyl caprate (C10) are assigned RI values of 800, 900, and 1000, respectively. The RI of a given metabolite is calculated according to the retention time relative to those of the FAMES species that elute immediately before and after it.

1. A RI calibration (RTC) file must be created for the first time of analysis on your system. A technical brief regarding this can be downloaded from the Agilent website (*MassHunter Quant Software: Incorporating Retention Index Results in Deconvoluted GC/MS Library Search Data*, https://www.agilent.com/cs/library/datasheets/public/MassHunter_Technical_Brief_v3.pdf). A detailed workflow for generating the RTC file is provided in **Note 16**. Refer to Subheadings 3.5 and 3.6 for the detail on the MassHunter Unknowns Analysis and Quantitation Analysis, respectively.
2. Once the RTC file is created for the specific analytical setup, the same file can be repeatedly used for different batches. However, RT must be updated for each batch to avoid errors due to slight RT shifts. The batch-specific retention time of each FAME can be obtained from the FAMES mix sample data file by MassHunter Qualitative Navigator. Open the data file of the FAMES mix sample in the batch, and display EIC for m/z 87.0 (go to Chromatogram menu from the main window, select Extract chromatogram, and select QC sample, Type

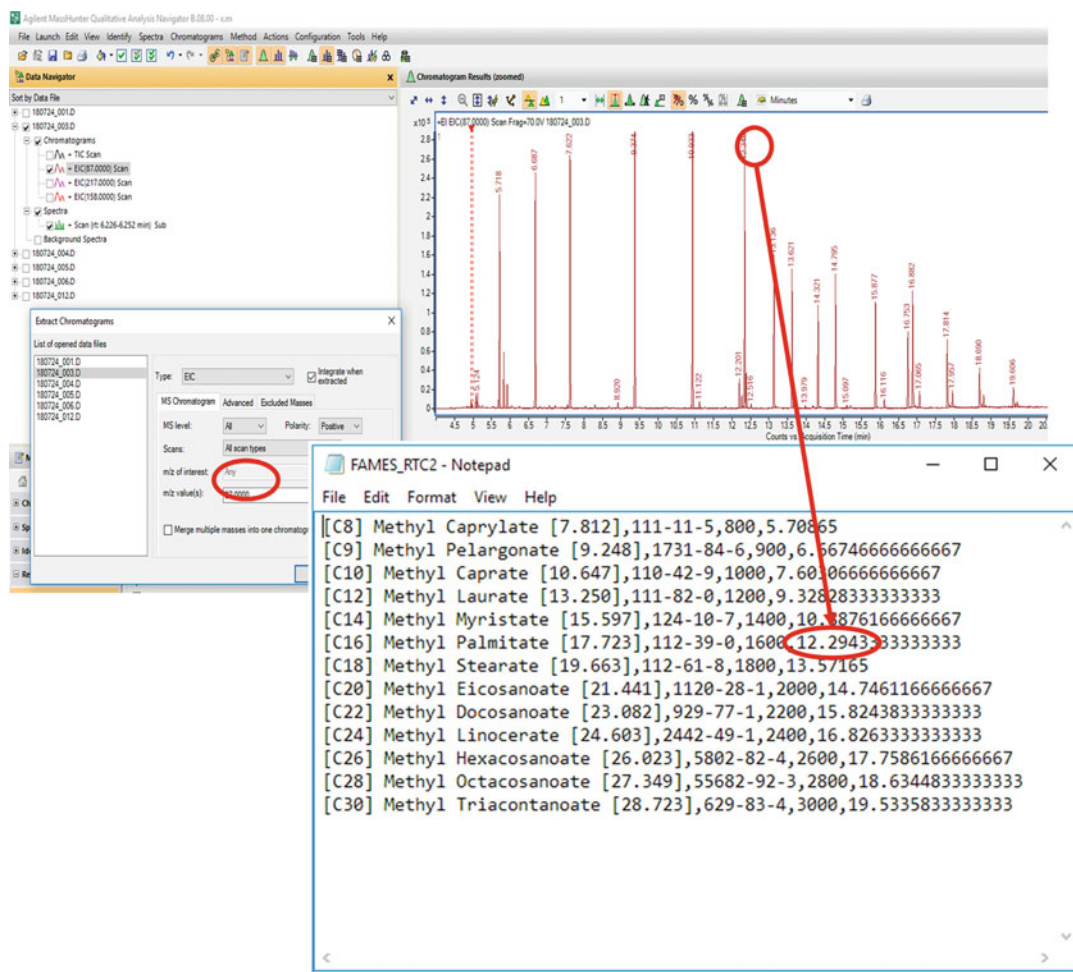


Fig. 3 Updating the retention time calibration file for each batch. An available *.RTC file can be updated for individual batches by editing the retention time of each FAME (lower panel). The batch-specific retention time can be obtained by reading the retention time of each FAME in MassHunter Qualitative Navigator. The FAME peaks are easily identified by the extracted ion chromatogram for $m/z = 87$

“EIC,” and m/z value(s) = “87”) since all the MSTFA derivatized FAME species generate an ion of m/z 87. Record the retention time of each FAME peak.

- Open the RTC file using excel or a text editor. Replace the retention time information for each FAME with the recorded values, and save the file as a batch-specific RTC file (Fig. 3).

3.5 Peak Annotation by MassHunter Unknowns Analysis Software

A batch-specific metabolite library is created in this step. Peaks in the reference sample are annotated by library search using Fiehn library to make a list of metabolites to be quantified in the batch. The idea is that the reference sample should contain all metabolites in the samples analyzed in the batch and the peaks annotated in that

sample are the ones to be quantified in the next step. Additionally, the resulting library contains precise retention time and high-resolution mass spectrum information, which greatly enhance the detection of metabolite peaks in the quantification step.

1. Open the “MassHunter Unknowns Analysis” software, and create a new analysis (File -> New Analysis; create this new batch file in the same folder as the *.D files are located).
2. Add data of the QC sample (Home -> Add samples; select *.D folder of the QC sample).
3. Go to Home->Method->Edit method. Change following in the Peak Detection tab:
 - (a) Peak Detection: select Deconvolution from the drop-down menu.
 - (b) In Peak filter: Exclude m/z: add 73:75,147:149. In the GC-MS protocol described here, the fragments with m/z 73,74,75,147,148, and 149 are ubiquitously generated from compounds carrying a trimethylsilyl moiety. It is beneficial to exclude them from subsequent analysis. Apply a height filter: Check the Absolute height box and add 2000. The height filter 2000 usually works well, but this value can be adjusted depending on the concentration of the samples.
 - (c) Keep default setting in the Deconvolution tab. Set the RT window size factor as 25,50,100,200 (*see Note 7*).
 - (d) In the Library Search tab, add library -> select Fiehn-2013.L file. Keep pre-search type: Normal; check Adjust Score; check Remove Duplicate hits; keep Pure Weight Factor: 0.7. In the Match factor section, uncheck Use RT Match. By selecting Adjust Score will give the closest match in the library. Selecting remove duplicate hits will help to minimize duplicates for a given target spectrum and returns a single library hit with the highest match score. Select the RT calibration file by clicking the Choose button, and select the RT calibration file created in Subheading 3.4.
 - (e) In the Compound Identification tab, set Max hit count to 1; Min match factor, 50; Min MZ, 60; Library search type, spectral search; and Multi-library Search Type, ALL (use this option in case multiple libraries are being used for peak search).
 - (f) In the Target Match tab, check the Within target RT window. Check Use compound name and Use CAS#. These parameters are related to the quantitation of targets by the quantitation analysis described in Subheading 3.6.

- (g) In the blank subtraction tab, uncheck Perform Blank subtraction.
 - (h) Click “apply to all samples” and close.
 - (i) Save the newly created Unknowns Analysis method for future use (Home -> Method -> Save Method).
4. Run peak detection and library search (Home -> Analyze -> Analyze All).
 5. Once the analysis is finished, display all hits (Home > Component Filter; highlight only “Hit”; remove “Best Hit”). Order the hits according to the RI difference (click “Delta RI” column name of the “Components” table), and delete results with RI difference <-150 or >150 . Display only the best hits (Home > Component Filter; highlight “Best Hit”).
 6. Manual curation of peak annotation and delete/correct according to the following criteria. For more details, *see* **Notes 8 to 15**.
 - (a) The RI should be close to the library data. Generally, delta RI should be within ± 50 till RI ~ 1500 , within ± 100 RI between 1500 and 2500, and within ± 150 in RI beyond 2500. The delta RI values of closely eluted compounds should be very similar especially when they are eluted in the same section of retention time divided by two FAMES. This is a more reliable criterion than the absolute delta RI values.
 - (b) Mass spectra should be similar to those in the library data. Ideally, more than three monoisotopic fragment matchings should be found with the library data excluding the fragment with m/z 73–75 and 147–149. However, a lower number of matching fragments can be allowed when only one or two prominent ions can be produced from the compound.
 7. Delete incorrect annotations (right-click the row; select “Delete components/hits”). To change annotation, right-click the row; select “Show Alternative hits.” This will open a new window showing alternative hits. The top panel shows the mass spectrum of the component, the middle panel shows the component and library match, and the lower panel shows the mass spectrum of the library compound. Manually inspect each alternative hit and select a compound that is more likely according to the criteria in Subheading 3.5, **step 6**. Press the “Set Best Hit” button to replace the alternative hit with the original annotation.
 8. Once all the satisfactory annotations are selected (Fig. 4), export the Best Hit annotation as a library (File -> Export, Export Component Table, Export from: All components/hits,

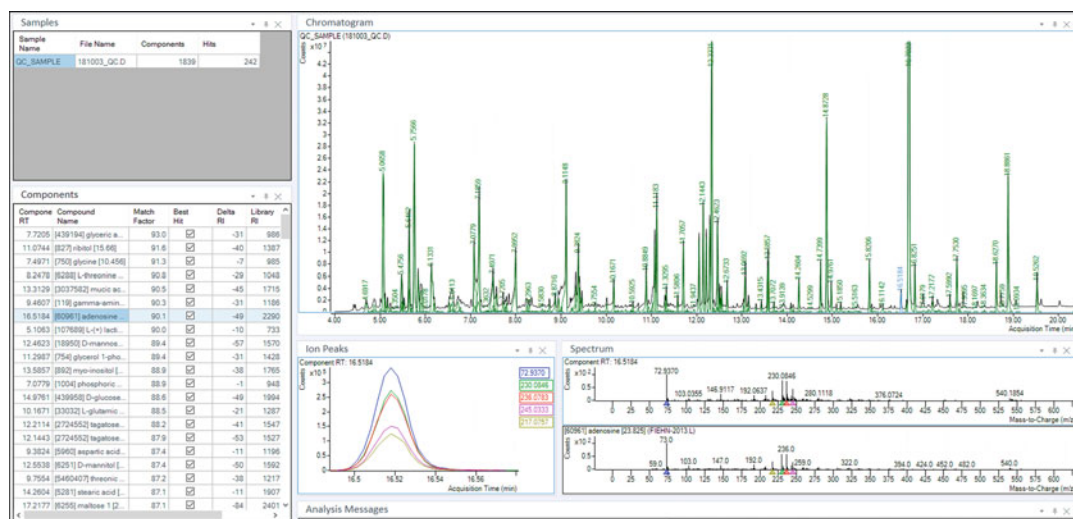


Fig. 4 Screenshot showing completed Unknowns Analysis. Component window (lower left) shows each manually curated identification of components with RT, compound name, Match Factor, whether it is the best hit or not, delta RI, and Library RI information. The Chromatogram window (upper right) shows the compounds identified by library search in green. The Ion peaks window (lower middle) shows the extracted ion chromatogram of each EI fragment contained in the annotated component. The upper and lower panels of the Spectrum window (lower right) show detected and library spectra to check fragment match

Export to: Library, no check on Automatically -> OK). Save the created library in the same folder as the *.D files. Save the annotation information also as a *.csv file (File -> Export, Export Component Table, Export from: All components/hits, Export to: CSV files, no check on Automatically -> OK).

9. Generate a report of the Unknowns Analysis to record the parameters used for peak annotation (Home-> Report -> Generate Report). Under the report method, choose New, check Selected Sample(s), and select Generate reports now. This will open up a new pop-up window. Right-click on the Templates area -> Add template and select the best-hits.template.xml file in MassHunter -> Report Templates -> Quant -> PDF-Reporting-> Unknowns. Report format can be selected between PDF, TEXT, and CSV. Save the method in the same folder with that of *.D folders and exit. The report file is shown in the Report method area. Click OK. An example of Unknowns Analysis with hits and generated reports is shown in Figs. 4 and 5.

3.6 Peak Quantification Using MassHunter Quantitative Analysis Software

The peaks annotated to metabolites are quantified in all samples in this step using batch-specific library created in Subheading 3.5 [8]. Due to the precise retention time and high-resolution mass spectrum in the batch-specific library, very strict parameters can be applied for library search, allowing the detection of identical peaks

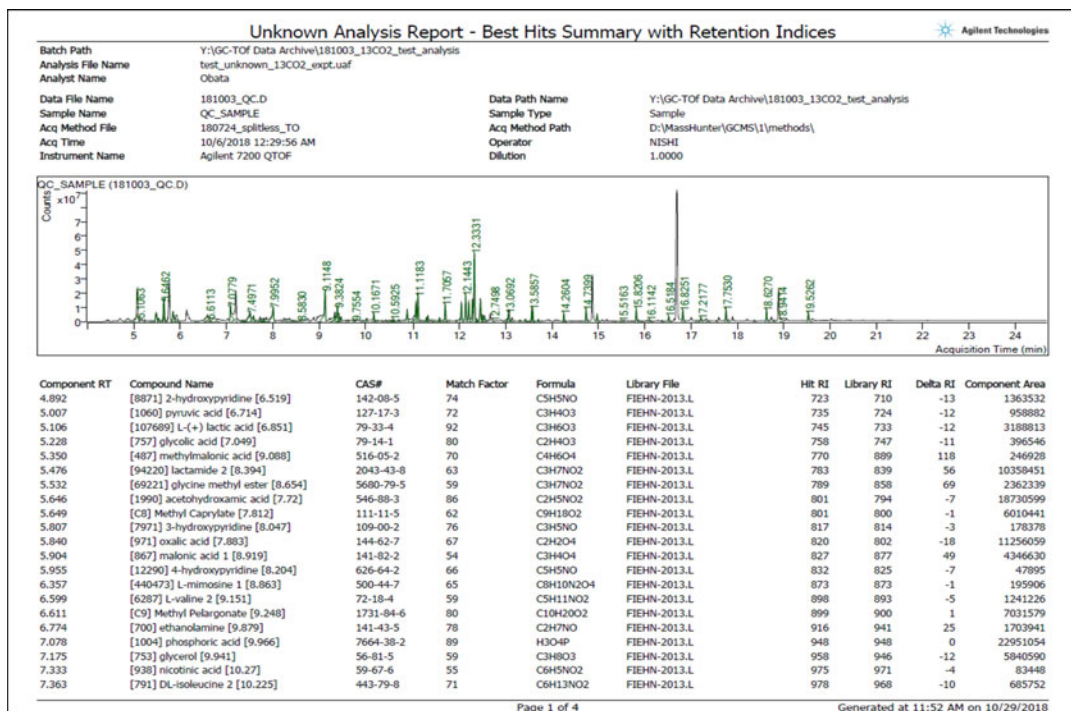


Fig. 5 An example of a best-hits report from Unknowns Analysis. The top part of the report shows information and metadata of the sample, including the name of the sample, name of the operator, sample acquisition date, and sample location. The middle panel shows the chromatogram of the components. The lower part indicates the identified compounds with RT, name, match factor, molecular formula, the spectral library used for identification, Hit RI, Library RI, Delta RI, and peak height of the component

from all samples in the batch. The peak height of the selected quantifier ion represents the relative abundance of the metabolite. This process is computationally demanding and takes a long time. It is recommended to store *.D files in a local drive. Also, consider determining suitable settings first with selected samples, and then import all samples for quantitative analysis of that batch.

1. Create a new batch in MassHunter Quantitative Analysis. Open the software and create New Batch in the same folder as the *.D files (Home -> Batch -> New Batch). Load all the blanks, reference samples, and sample files to be included in the batch. Select the reference sample first in the batch to be used to generate the quantitation method (as this file is used in Unknowns Analysis to identify and annotate the peaks).
2. Create a new quantitation method based on the library file created by the Unknowns Analysis in Subheading 3.5 (Method -> New -> New method from file and choose the library file created by the Unknowns Analysis in Subheading 3.5). On the “Method from Library” window, change the following settings, and then press OK, Library, leave as it is (library file

created by the Unknowns Analysis in Subheading 3.5); Targets, All compounds; Spectra, Create targets per spectrum in the library; Target ion, Weighted; and Retention time calibration, Use RT Calibration; choose the *.RTC file generated in Subheading 3.4 and used for the library search in Unknowns Analysis; uncheck “Update RT from sample data”; for Qualifiers, uncheck Sum qualifier(s); and set “Number of Qualifiers to add” to 5.

3. Go to the View menu and click Panes to open the Methods Task Panel. Set deconvolution parameters (Method Tasks > Workflow > Target Deconvolution) as follows: Setup Reference Library, select the “Obtain reference library from lookup library” option, and select the library generated previously in Subheading 3.5. Create reference library at: select the folder containing *.D files. Press OK.
4. Create a new library method (Method Tasks > Library Method > New). Follow the setting used in the peak detection method for the Unknowns Analysis (Subheading 3.5, step 3). Go to the “Library Search” tab, and change the library to the reference library just created in step 3. Set RT calibration file to the *.rtc file used for Unknowns Analysis. In Match factor tab, check “Use RT Match”; select “Gaussian” with standard deviation of 1 sec; and select RT mismatch penalty “Multiplicative”. In Blank Subtraction tab, check off “Perform Blank Subtraction”. Save the library method. Then, check “Show reference spectrum,” “Show override spectrum,” “Show match scores” and check off “Deconvoluted” and press OK.
5. Adjust extraction MZ window (Method Tasks > Method Setup Tasks > Compound Setup; show columns “MZ Extraction Window Units,” “Extract Left m/z,” and “Extract Right m/z”; set MZ Extraction Window unit “PPM,” Left m/z “200,” and Right m/z “200”).
6. In Retention Time Setup, adjust Left and Right RT Delta to 0.05 minutes.
7. Select appropriate quantifier and qualifier ions in Qualifier Setup. Add the Ion Polarity column to the Qualifier, select Positive, and fill it down to avoid an error due to the lack of ion polarity information. Remove ions with m/z 70–79 and 147–149. If these ions are selected as the quantifier ion, swap it with one of the qualifier ions. Set “Ion Polarity” to “Positive” when the quantifier is swapped. In case the automatically selected ion is not the one representing the abundance of the metabolite due to the reasons including too high or too low abundance and intensity affected by co-eluting compounds, select a quantifier ion free from these effects.

8. Adjust global setting (Method Tasks > Method Setup Tasks > Globals Setup): Check the Bracketing Type “Overlapped” and “Ignore Peaks Not Found.” Library Method and Reference Library should be updated from the Deconvolution Setup. Add Non Reference Window value 200 minutes, no Reference Pattern Library, and Reference Window 80 Percent; uncheck all other settings.
9. Save the method file as Quantification Method in the same folder as the *.D files. Click “Exit” and analyze all samples (do not need to convert to DTA files).
10. Manually curate the quantification result and refine the method.
 - (a) Click the “Display Multiple Compounds/Samples in Batch Table” button. Check if there are compounds with many missing data and highly varying values.
 - (b) Check the precise mass of the quantifier ion if it is different from the library value. Consider changing the product ion m/z to detect. Sometimes a very prominent fragment peak is not detected by Quantitative Analysis software. In most cases, this is due to a slight difference of precise mass in the range of three decimal points between Unknowns Analysis and Quantitative Analysis software. The difference of m/z 0.010 sometimes compromises peak detection. This can be solved by manually obtaining the precise mass values using MassHunter Qualitative Navigator. The target peak is identified by the retention time and the EIC of nominal mass in the Qualitative Navigator. Precise mass can be checked by showing the mass spectrum of the peak. When the correct precise mass is used for the quantitative method, narrower MZ ranges (e.g., 100 ppm for both directions) tend to give better results. Alternatively, if a minor ion was selected as the quantifier ion, consider swapping it manually to a major one.
 - (c) Check the retention time if some samples showed retention time shift. Consider setting a wider RT window. In case the target compound elutes closely to other compounds with a similar ion, consider a narrower RT window.
 - (d) Click the “Display Single Compound/Sample in Batch Table” button.
 - (e) Check that each compound has at least one qualifier ion which is in an acceptable intensity ratio with quantifier ion for most of the samples.

If the compound has only one prominent ion, it may have no qualifier ion in a good range. Other compounds with no good qualifier ion may be due to co-eluted

compounds, low abundance, or too high abundance exceeding detection limit (overloading). Consider changing the quantifier ion or removing the compound from the analysis.

- (f) Change Quantification Method and run analysis. Repeat the steps above till all the samples show acceptable results.
11. Repeat method refinement till all metabolites in all samples show acceptable results. When it is difficult to gain consistent values among samples, check the raw chromatograms of selected samples by MassHunter Qualitative Navigator whether it is due to the variation of peak heights, detection of different peaks, or too low/high peak intensities. In the latter two cases, repeat method refinement or consider removing the peak from the analysis.
12. Export the data. Click the “Display Multiple Compounds/Samples in Batch Table” button. Right-click a compound name and select “Add/Remove Columns” and display only “Height.” Right-click somewhere on the table, and select “Export Table” to save the peak height as a .csv table. Repeat this for “Product ion” and “RT.” Combine all of them into a single Excel workbook as the result file. Examine the product ion and RT data, and make sure that the same product ion was selected as the quantifier ion for a particular metabolite across all the samples. These data, together with the report from Unknowns Analysis, can be used to assess the quality of peak annotation and recommended to submit together with the results when they are published.

3.7 Initial Normalization, Filtering, and Statistical Analysis

1. To investigate whether all profiles are acquired at the same analytical precision, analyze the peak heights of the internal standard (ribitol). The peak height of the internal standard is log-transformed and plotted as a scatter plot. If the peak height of ribitol is nearly constant (barring some random variance, max allowable percentage of cumulative variation is 20%), the constant condition can be validated throughout the experimental procedure. In case some samples show abnormal values, check the chromatogram, and remove them from further analysis if the samples have critical problems such as injection errors or failure in derivatization.
2. To remove the peaks which are not derived from biological samples, the peak height values of each metabolite in the blank sample are subtracted from those in each sample. If multiple blank samples were analyzed in the batch, the average peak height in the blank samples can be subtracted. In case the resulting values are very small (i.e., smaller than 20% of blank

values) or negative in most of the samples, discard the annotation since it unlikely reflects metabolite levels in the samples.

3. To normalize the technical errors, the height values of each metabolite peak are divided by that of internal standard, ribitol, within each sample. A second normalization is performed using sample weight to normalize the differences derived from the deviation in the amount of material used for the extraction. The values of each metabolite are divided by the weight of that sample. This will bring the normalized peak height values per unit mass of the sample. This value represents the relative level of a metabolite which can be used to compare the abundance of each metabolite between samples. Note that the values cannot be used to analyze across metabolites.
4. Further normalization by the mean metabolite level in control samples or by the median metabolite level in all samples is useful to make the data more understandable depending on the experimental design. Additional log transformation should be considered to achieve normal distribution of the data and to give equal impact on increased and decreased metabolites.
5. Data of known technical artifacts from column bleeding, reagents, and internal standard including FAMEs should be removed. Some metabolites appear as multiple peaks, and redundant data should be deleted at this point (*see* **Note 8**).
6. The metabolite profile data matrix after normalization and filtering can be further used for statistical analyses. Multivariate statistical analysis methods can be applied to group metabolites and samples which showed similar profiles. Principal component analysis (PCA) [28] is a widely used unsupervised dimensionality reduction method. It ranks the principal components based on the variance in the features (here metabolites). But PCA is performed without considering the sample trait or experimental group under investigation. To detect the similarities and differences between the samples based on the overall metabolite profile, methods such as hierarchical clustering analysis (HCA) [29] can be applied. HCA will cluster samples that have overall similar profiles and thus can be used to discriminate samples without the prior knowledge of sample class. Using the combination of variable importance in the projection (VIP) and partial least square-discriminate analysis (PLS-DA) [30] discriminative power of each metabolite can be quantitatively estimated. This regression and classification method relies on the class label of the sample (here experimental group/phenotypic trait) and identifies important individual features that have the greatest effect on the classification [31, 32].

7. Further, univariate analysis can be used to detect metabolites that are statistically significantly changed across different experimental groups. As the number of metabolites to be analyzed in this protocol is typically dozens to 100 and multiple metabolites are not detected for all metabolic pathways, a classical analysis of individual metabolites is often useful. If data has two experimental groups (e.g., control and treatment), fold change analysis can be performed to compute the $\log_2$ fold change values of each metabolite. By applying a t-test, a raw p-value and FDR corrected p-value (using either Holm, Hochberg, Bonferroni, or BH (Benjamini and Hochberg)) should be computed. For multi-group analysis, analysis of variance (ANOVA) and associated post hoc tests can be performed to identify significant metabolites that follow a given pattern.
8. A short compiled list of software tools and workflow available for analysis of metabolomics data is provided in Table 1. An excellent repository for different analysis tools is provided by OMICtools [33]: a website of manually curated tools for the analysis of different omics data (both commercial and open-source). It provides basic information on functionality, computer skills, licensing, programming language, and interface, but it does not contain information that a user will be required or a decision tree to decide which tool to use and the data format for as input in the tools.

4 Notes

1. It is advisable to use tubes produced by Eppendorf as the inner surface of these tubes is inert and ideal for sample preparation and long-term storage. Tubes from other makers can be used, but some cause huge background and artifact peaks derived from compounds from the inner surface. An initial test is strongly recommended before using a new product.
2. Metabolite contents in plant samples are typically high due to low water contents compared with animal and microbial samples. Metabolite amount is usually enough with less than 10 mg of fresh plant material. However, taking 50–100 mg of material ensure that the material is weighed accurately to minimize sample-to-sample variation.
3. Ribitol concentration should be optimized according to the GC system and metabolite amount in the sample. When the metabolite contents in the starting material are low and the internal standard peak is overloading, the normalization by ribitol in Subheading 3.7, **step 4** does not work to cancel the technical errors.

Table 1
Software tools and workflow for the statistical analysis of metabolomics data

Tool/workflow	Data type	Software type	Website	Short description/utility
Metabolomics	MS, NMR	R package	https://cran.r-project.org/web/packages/metabolomics/	A collection of functions for statistical analysis of metabolomics data
MetaboLyzer	MS, NMR	Command line	https://sites.google.com/a/georgetown.edu/fornace-lab-informatics/home/metabolizer	A python-based statistical analysis toolset for metabolomics
Muma	MS, NMR	R package	https://cran.r-project.org/web/packages/muma/	R package for preprocessing of metabolomics data. This package includes functions for univariate and multivariate analysis of metabolomics data
Ropls	MS, NMR	R package	https://www.bioconductor.org/packages/release/bioc/html/ropls.html	This package implements the PCA, PLS-DA, and OPLS-DA algorithms for multivariate analysis of omics data
Workflow4metabolomics	LC-MS, GC-MS	Web-based	http://workflow4metabolomics.org	A full web-based suite for LC/MS, GC/MS, and NMR analysis pipeline using Galaxy framework for preprocessing, normalization, and statistical analysis of metabolomics data
XC-MS online	LC-MS, GC-MS	Web-based	https://xcmsonline.scripps.edu/landing_page.php?pgcontent=mainPage	A web-based implementation of xcms R package for data analysis of metabolomics experiments
MetaboAnalyst	LC-MS, GC-MS	Web-based	http://www.metaboanalyst.ca	Web-based tool for statistical analysis of metabolomics data

4. Based on the sensitivity of the analytical instrument, the amount of extract to be dried should be optimized. Although the dynamic range of GC-MS instruments is fairly wide, it is advisable to determine a suitable amount of metabolite extract, especially prior to large-scale analysis. The metabolite content can also be adjusted by the volume of derivatization reagents and by changing the split ratio although these are less flexible.
5. In the presence of water, MSTFA derivatization efficiency significantly reduces. Especially in a humid region, the use of a glove box filled with dry nitrogen should be considered. Moreover, introducing water in the GC-MS system is not advisable as this will ruin the column, reduce the sensitivity of the detector, and might cause oxidation of the filament. It is beneficial to use anhydrous reagents for the derivatization of metabolite extract. This will minimize water contamination during derivatization and keep the GC-MS in excellent condition.
6. Before running all samples in the batch, the split ratio should be optimized according to the expected metabolite concentration for the sample under consideration. We usually test splitless and split ratios of 1:10, 1:100, and 1:500 with the pooled reference sample as a preliminary test and then decide which ratio works best. In many plant materials, sugars, including glucose, fructose, and sucrose, are highly accumulated, and more than one split ratio is required to determine all metabolites. In this case, the same sample set can be run repeatedly for multiple split ratios.
7. The window size factor is a dimensionless scale of correlation window for grouping ion peaks (equivalent to Resolution in AMDIS). Selecting smaller value separates closely spaced peaks, finds more components, but runs longer. It is advisable to use multiple values to cover all kinds of peaks with minimum manual intervention.
8. Especially for sugars and polyols, check the alternative hits. If there is another hit with close RI, select that as the best hit. Molecules with TMS 1 and 2 (e.g., L-Valine 1 and L-Valine 2) are the different forms of the same molecule; both can be selected. Both should behave in the same way, and one of them can be selected when the final result table is created. If the same annotation appears multiple times, only one of them is correct. If the delta RI is too high, check the alternative hits, and close library retention time to neighboring compounds. The library RI of the alternative hits cannot be seen in the Unknowns Analysis but can be found by Library Editor.
9. Norvaline is most likely valine; check alternative hits. Similarly, leucine, isoleucine, and norleucine are often mis-annotated.

10. Usually the most abundant sugars in plant samples are fructose, glucose, and sucrose; they are often misannotated as psicose, tagatose, and others.
11. Alanine shows only one or two fragments and tends to show a low match factor. Check alternative hits for the components which have RI close to 777 (may need to check the hits other than the best hits).
12. Fumarate tends to be mis-annotated as maleate; check alternative hits.
13. Ribitol is sometimes mis-annotated as other polyols or sugars. This peak can be easily distinguished by the intensity since the internal standard shows a large peak height, which should be consistent among samples.
14. Citric acid can be annotated as isocitric acid. Both are possible and most likely show mixed peaks. Select citric acid and calculate later using specific fragments to isocitric acid (m/z 245.0333).
15. Minor sugars, disaccharides, and polyols are very hard to annotate with the library data due to their close retention times and almost identical mass spectra. Keep a record on your lab book to remember that the annotation might not be correct. The annotation should be checked using standard compounds and backup extracts.
16. Detailed workflow for generating new RTC file:
 - (a) Acquire GC-MS profile of a sample containing FAMES standard mix. The blank sample (MSTFA with FAMES mix) included in Subheading 3.2, **step 2** is suitable for this purpose.
 - (b) Using MassHunter Unknowns Analysis, deconvolute and search the spectra of components in FAMES mix sample against Fiehn 2013 library. Retain only the FAME hits, and use the export function in MassHunter Unknowns Analysis to create a user library in *.xml format. *See* Subheading 3.5 for more details.
 - (c) Create a new batch file in MassHunter Quantitative Analysis containing only the FAMES mix sample (Fig. 6a; *see* Subheading 3.6 for detail). Create a new quantitation method in the Method section: New > New Method from Acquired Scan Data with Library Search. This will open up the library method editor. Go to the Libraries on the Library Search tab, and change the library file to the FAMES library newly created by Unknowns Analysis at **step b**. This step creates a quantitation method file for the identification of FAME peaks. Refine the quantitation parameters as described in Subheading 3.6 until all

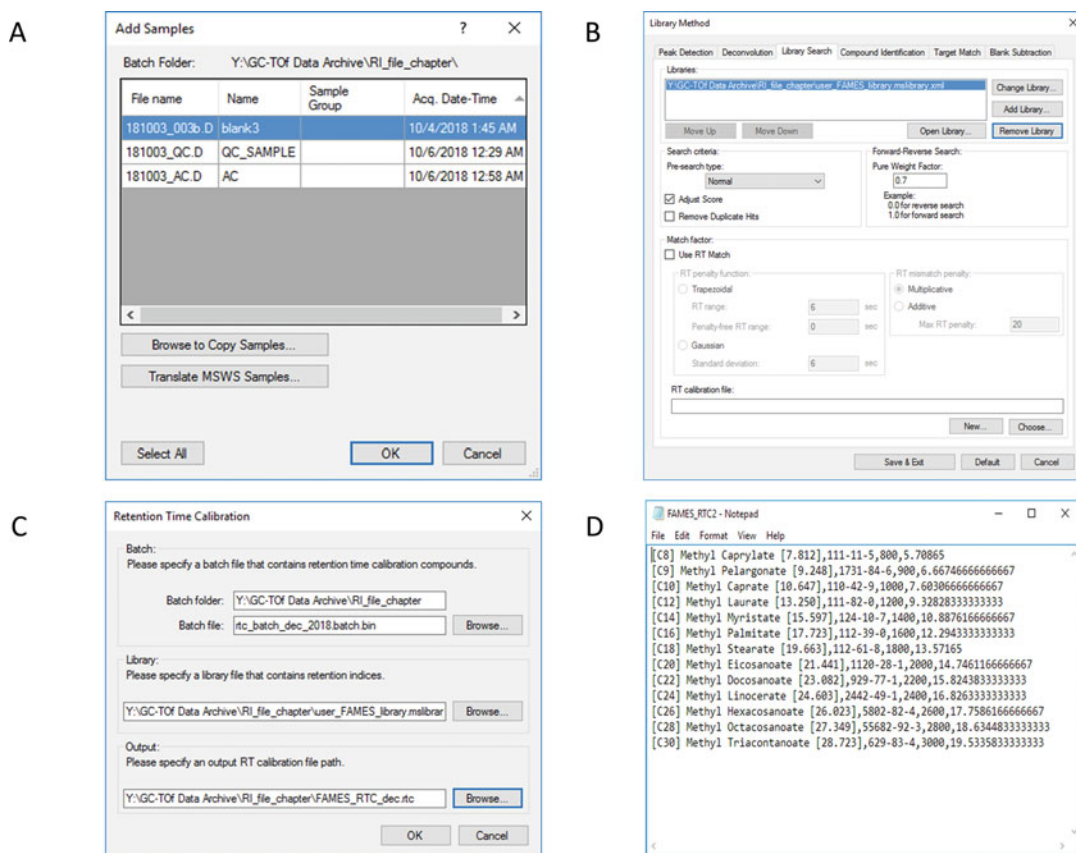


Fig. 6 Steps for generating initial RT calibration (*.rtc) file. (a) Add the blank sample containing the FAMES mix to the batch in MassHunter Quantitative Analysis software. (b) Library Search panel in the MassHunter Quantitative Library Method editor, where FAMES library generated by the Unknowns Analysis is loaded. (c) Retention Time Calibration window to define the files required to generate RTC files. (d) Example of a *.RTC file. This can be edited by Excel or a text editor to update

FAME peaks are correctly detected. Save the quantitation method as a new MassHunter Quantitative Analysis method file.

- (d) The RTC file is created in the Library Method editor in MassHunter Quantitative Analysis software (Method tab > Library > Library Method). Choose the method file created in **step c** and click Edit. This opens up the Library Method editor (Fig. 6b). Go to the Library Search tab and select the library created in **step b**. Then go to the RT calibration file option at the bottom of the window and click New (Fig. 6b). This opens up the Retention Time Calibration window (Fig. 6c). Browse to select the Quantitation batch file created in **step c**. In the Library section, specify the library file as the one created in **step c**. The output path and the name of the RTC file are specified in

the Output section, and then click OK. This creates a new RTC file that is a simple text file containing the names, CAS numbers, nominal RI values, and the retention time in minutes of the FAME species (Fig. 6d).

Acknowledgments

This study was supported by the National Science Foundation/EPSCoR RII Track-2 FEC (Award #1736192, to N.A. and T.O.), the National Science Foundation/EPSCoR RII Track-1 (Award #1557417, to N.W. and T.O.), and the University of Nebraska-Lincoln Faculty Startup Grant (to T.O.).

References

1. Tuberosa R (2012) Phenotyping for drought tolerance of crops in the genomics era. *Front Physiol* 3:347
2. Furbank RT, Tester M (2011) Phenomics - technologies to relieve the phenotyping bottleneck. *Trends Plant Sci* 16(12):635–644
3. Harrigan GG, Martino-Catt S, Glenn KC (2007) Metabolomics, metabolic diversity and genetic variation in crops. *Metabolomics* 3(3): 259–272
4. Neilson EH, Edwanrds AM, Blomsleedt CK et al (2015) Utilization of a high-throughput shoot imaging system to examine the dynamic phenotypic responses of a C4 cereal crop plant to nitrogen and water deficiency over time. *J Exp Bot* 66(7):1817–1832
5. Patti GJ, Yanes O, Siuzdak G (2012) Innovation: metabolomics: the apogee of the omics trilogy. *Nat Rev Mol Cell Biol* 13(4):263–269
6. Steinfath M, Strehmel N, Peters R et al (2010) Discovering plant metabolic biomarkers for phenotype prediction using an untargeted approach. *Plant Biotech J* 8(8):900–911
7. Zhang Y, Zhang Y (2007) Formation and reduction of acrylamide in Maillard reaction: a review based on the current state of knowledge. *Crit Rev Food Sci Nutr* 47(5):521–542
8. Fernie AR, Schauer N (2009) Metabolomics-assisted breeding: a viable option for crop improvement? *Trends Genet* 25(1):39–48
9. Dunn WB, Bailey NJC, Johnson HE (2005) Measuring the metabolome: current analytical technologies. *Analyst* 130(5):606–625
10. Obata T, Fernie AR (2012) The use of metabolomics to dissect plant responses to abiotic stresses. *Cell Molec Life Sci* 69(19): 3225–3243
11. Hoker J, Obersteine F, Bönisch H et al (2015) Comparison of GC/time-of-flight MS with GC/quadrupole MS for halocarbon trace gas analysis. *Atmosph Measur Technol* 8(5): 2195–2206
12. Dunn WB, Broadhurst D, Begley P et al (2011) Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat Protoc* 6(7):1060–1083
13. Lisec J, Schauer N, Kopka J et al (2006) Gas chromatography-mass spectrometry-based metabolite profiling in plants. *Nat Protoc* 1(1):387–396
14. Dekoning W, Vandam K (1992) A method for the determination of changes of glycolytic metabolites in yeast on a subsecond time scale using extraction at neutral pH. *Anal Biochem* 204(1):118–123
15. Gonzalez B, Francois J, Renaud M (1997) A rapid and reliable method for metabolite extraction in yeast using boiling buffered ethanol. *Yeast* 13(14):1347–1355
16. Hancock R (1958) The intracellular amino acids of *Staphylococcus aureus* - release and analysis. *Biochim Biophys Acta* 28(2):402–412
17. Gale EF (1947) The assimilation of amino-acids by bacteria: 1. The passage of certain amino-acids across the cell wall and their concentration in the internal environment of *Streptococcus faecalis*. *Microbiology* 1(1):53–76
18. Fuerst R, Wagner RP (1957) An analysis of the free intracellular amino acids of certain strains of *Neurospora*. *Arch Biochem Biophys* 70(2): 311–326
19. Maharjan RP, Ferenci T (2003) Global metabolite analysis: the influence of extraction

- methodology on metabolome profiles of *Escherichia coli*. *Anal Biochem* 313(1): 145–154
20. Rabinowitz JD, Kimball E (2007) Acidic acetonitrile for cellular metabolome extraction from *Escherichia coli*. *Anal Chem* 79(16): 6167–6173
 21. Chen SL, Hoene M, Li J et al (2013) Simultaneous extraction of metabolome and lipidome with methyl tert-butyl ether from a single small tissue sample for ultra-high performance liquid chromatography/mass spectrometry. *J Chromat A* 1298:9–16
 22. Kind T, Wohlgemuth G, Lee DY et al (2009) FiehnLib: mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. *Anal Chem* 81(24): 10038–10048
 23. Kanani H, Chrysanthopoulos PK, Klapa MI (2008) Standardizing GC-MS metabolomics. *J Chromat B Anal Technol Biom Life Sci* 871(2):191–201
 24. Jonsson P, Gullberg J, Nordström A et al (2004) A strategy for identifying differences in large series of metabolomic samples analyzed by GC/MS. *Anal Chem* 76(6):1738–1745
 25. Xu FG, Zou L, Ong CN (2010) Experiment-originated variations, and multi-peak and multi-origination phenomena in derivatization-based GC-MS metabolomics. *TrAC Trends Anal Chem* 29(3):269–280
 26. Moros G, Chatziioannou AG, Gika HG et al (2017) Investigation of the derivatization conditions for GC-MS metabolomics of biological samples. *Bioanalysis* 9(1):53–65
 27. Fang ML, Ivanisevic J, Benton HP et al (2015) Thermal degradation of small molecules: a global metabolomic investigation. *Anal Chem* 87(21):10935–10941
 28. Raychaudhuri S, Stuart JM, Altman RB (2000) Principal components analysis to summarize microarray experiments: application to sporulation time series. *Pac Symp Biocomput*:455–466
 29. Eisen MB, Spellman PT, Brown PO et al (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* 95(25):14863–14868
 30. Worley B, Halouska S, Powers R (2013) Utilities for quantifying separation in PCA/PLS-DA scores plots. *Anal Biochem* 433(2): 102–104
 31. Ch HW, Broeckling CD, Helmus R et al (2008) Discovery of metabolite features for the modelling and analysis of high-resolution NMR spectra. *Internet J Data Min Bioinform* 2(2):176–192
 32. Gromski PS, Muhamadali H, Ellis DI et al (2015) A tutorial review: metabolomics and partial least squares-discriminant analysis - a marriage of convenience or a shotgun wedding. *Anal Chim Acta* 879:10–23
 33. Henry VJ, Bandrowski AE, Pepi AS et al (2014) OMICtools: an informative directory for multi-omic data analysis. *Database* (Oxford)



Gene Co-expression Network Analysis and Linking Modules to Phenotyping Response in Plants

Qian Du, Malachy T. Campbell, Huihui Yu, Kan Liu, Harkamal Walia, Qi Zhang, and Chi Zhang

Abstract

Environmental factors, including different stresses, can have an impact on the expression of genes and subsequently the phenotype and development of plants. Since a large number of genes are involved in response to the perturbation of the environment, identifying groups of co-expressed genes is meaningful. The gene co-expression network models can be used for the exploration, interpretation, and identification of genes responding to environmental changes. Once a gene co-expression network is constructed, one can determine gene modules and the association of gene modules to the phenotypic response. To link modules to phenotype, one approach is to find the correlated eigengenes of given modules or to integrate all eigengenes in regularized linear model. This manuscript describes the method from construction of co-expression network, module discovery, association between modules and phenotypic data, and finally to annotation/visualization.

Key words Gene co-expression network, Module discovery, Data integration, Association between co-expression network and phenotyping data

1 Introduction

Study on the link between genotype and phenotype is challenging because it involves numerous changes in biological pathways and processes. Since many genes are involved in the response to the perturbation of the environment, traditional approaches that examine one or a few genes may fail to capture and characterize the complex responses at the molecular level. Therefore, identifying functional gene clusters would be more meaningful than searching for a single gene. With the advent of next-generation sequencing technology, transcriptional responses to environmental stimuli can be examined at a genome-wide level and provide a comprehensive understanding of the complex processes underlying environmental adaptation and stress responses. In this manuscript, we discuss a

method to link a gene co-expression network constructed based on RNA-seq data to phenotyping data.

Environmental factors, including different stresses, can have an impact on the expression of genes and subsequently the phenotype and development of plants. High-throughput manner and low cost of sequencing technologies enabled the biologists generating a plenty of data for quantifying the relationship among genes that are involved in the response to environmental changes. These genes and interactions could be modeled as nodes and edges in a network model. Genes are represented as nodes in a network, and transcriptional interaction is the connection between nodes [1–3]. Specifically, the gene co-expression network models have been used for the exploration, interpretation, and visualization of the relationship among genes in a wide range of biological applications, including disease-gene association [4], identification of genes responding to environment changes, tissue-specific gene identification [5], and functional gene annotation [6]. For example, RNA sequencing data provide valuable information on gene expression across different conditions, time points, tissues, or genotypes. In co-expression network analysis, genes with similar expression pattern are grouped together, with the underlying rationale being “guilt by association.” This extensively validated principle states that transcriptionally coordinated genes are often functionally related. Gene co-expression networks can also be used to study the association between genotype and phenotypes, such as identifying functional eQTLs [7] and studying gene-phenotype association [8] after being combined with other biological data in various downstream analysis.

Many construction tools and analysis tools were developed for gene co-expression network. WGCNA [9] is based on the marginal correlation of the gene pairs. GeneTS [10] and BicMix [11] built Gaussian graphical models (GGM) among gene pairs based on their conditional dependence under the multivariate Gaussian assumption. CorSig [12] can estimate the statistical significance of correlation for co-expression network analysis. The R package, PCIT, can construct a weighted gene co-expression network based on partial correlation and information theory approaches [13]. FastGCN can apply GPU to accelerate gene co-expression network construction [14]. VCNet can build a vector-based gene co-expression from RNA-seq data [15].

Once co-expression modules are identified, one can determine which modules are associated with the phenotypic response and which biological processes genes in the same module are involved. To link modules to phenotype, one approach is to calculate the correlation between physiological traits and eigengenes of given modules, which are defined as the first principal component of a specific module [16], and the other way is to integrate all eigengenes together to fit a linear model to the phenotyping data. The

genes in the selected modules that are associated with the phenotype data are candidates to respond to the environmental changes. These genes can undergo Gene Ontology term enrichment analysis and/or have visualization.

2 Method

2.1 Construction of Co-expression Network

In a gene co-expression network, one gene is represented as one node, and two genes are connected by an edge if the similarity of their expression profiles is high (*see Note 1*). The data used for constructing a gene co-expression network is represented as a $m \times n$ matrix, i.e., m genes for n samples (*see Notes 2 and 3*). Network construction has two steps: calculating co-expression measure and selecting significance threshold [17, 18].

For co-expression measure, a gene expression similarity score between each pair of gene expression profiles needs to be defined. The expression values can be represented as a vector, so evaluating similarity score between a pair of genes is similar to evaluating measure for two vectors of numbers. Usually, Pearson's correlation coefficient, which can be calculated in R with `cor(expression_matrix, methods = "pearson")`, is used to construct gene co-expression networks (*see Note 4*).

To construct a unweighted gene co-expression network, for all gene expression similarity scores, a cutoff is selected to determine the co-expressed genes [19]. For example, one can use the top 0.5% correlated pairs [20] (*see Note 5*). The WGCNA method uses soft thresholding, which outputs a weighted gene co-expression network after a power transformation of the raw similarity matrix. It constructs co-expression network using several thresholds and chooses the threshold which produces a network with scale-free topology.

2.2 Module Discovery

Many evidences have already showed that genes grouped together in the same module of a network work on a certain function, and hence, we tried to discover the functional modules from the gene co-expression network by identifying the intensively connected subgraphs, i.e., modules. One module, a group of co-expressed genes, can be identified using clustering algorithms [21]. Clustering is used in grouping genes which have similar expression patterns in several samples to build groups of co-expressed genes (*see Note 6*). The R package, WGCNA [9], identifies functional modules in our constructed co-expression network using hierarchical clustering. For using the R functions, `cutreeDynamic()` and `blockwiseModules()`, their parameters need to be tuned carefully for each dataset, and the parameter of `minModuleSize` is recommended to set as 5 to get the reasonable sizes of modules (*see Note 7*).

2.3 Link Phenotyping Data to Gene Modules

Due to multiple genes in one module, representative genes of the given module need to be defined, and usually, one can define an eigengene that is the first principal component of a specific module [16]. To link modules to phenotype, there are at least two types of approaches: one is to select associated modules based on the high correlation between physiological traits and eigengenes of the given module (*see Note 8*), and the other is to integrate all eigengenes together to fit a linear model with the phenotyping data as the response. To generate a given type of phenotyping data, many genes or pathways work together to respond to the environmental perturbation, and their contribution can be described by a linear combination, described by the following equation:

$$y = \sum_{k=1}^K w_k g_k + \varepsilon,$$

where y is the phenotyping data vector, the response of the plant, and g_k is the eigengene of module k , and w_k is weight of one eigengene contributing to the phenotypic response (*see Note 9*).

2.4 Gene Function Annotation

For modules and genes in modules that are linked to the phenotypic response, one can use enrichment analysis to annotate their functions. Functional enrichment analysis is a method to identify the overrepresentation of functional categories in a group of genes. If the co-expressed genes have a functional relationship, enriched functions of genes can be assigned to some poorly annotated genes in the module. This approach is known as guilt by association. The most popular way for enrichment analysis is to identify the enriched GO terms by using tools, such as GO::TermFinder [22] and Panther (<http://pantherdb.org/>). The P -value of enrichment can be calculated with hypergeometric distribution and further adjusted with Bonferroni correction for multiple testing. The cutoff of adjusted P -value can be 0.05 (*see Note 10*).

2.5 Visualization of Modules and Phenotyping Data

It is important to visualize the link between gene models and phenotyping data vectors. Cytoscape [23] can be used to draw the co-expression network of each module. In each module, genes with the same function can be displayed with the same color. The edge between two genes is shown if their correlation is statistically significant, i.e., adjusted P -value < 0.01 to reject the null hypothesis. The R function, *cor.test()*, will be applied for this test.

3 Materials

3.1 Databases of Gene Co-expression Networks in Plants

- ATTED II, <http://atted.jp/> [24].
- CoP, <http://webs2.kazusa.or.jp/kagiana/cop0911/> [25].
- GeneCAT, <http://genecat.mpg.de/cgi-bin/Ainitiator.py> [25].

- PlaNet, <http://aranet.mpimp-golm.mpg.de/> [26].
- ComPIEx, <http://complex.plantgenie.or> [27].
- AtGGM2014, <http://dinesh-kumarlab.genomecenter.ucdavis.edu/downloads.htm> [28].
- GeNET, bengi.cs.mun.ca/gene [29].
- VTCdb, <http://vtcdb.adelaide.edu.au/home.aspx> [30].
- OryzaExpress, <http://riceball.lab.nig.ac.jp/oryzaexpress/> [31].
- RiceAntherNet, <https://www.cpib.ac.uk/anther/riceindex.html> [32].

4 Notes

1. The node degree is identified as the number of connections of a specific node with other nodes.
2. Sequencing depth cutoff thresholds for RNA-seq data are usually selected arbitrarily. Several studies have used a cutoff of 10 million reads per sample. It is suggested that co-expression networks built using this cutoff will have a similar quality to microarray co-expression networks if constructed using the same sample size [18, 33]. Therefore, some genes will be discarded if their normalized read counts < the cutoff, such as 15, for all samples or all read counts have variable of coefficient < 0.3. Usually for a genome, 4000–6000 genes will be kept for the gene co-expression network construction.
3. Twenty samples at minimum have been suggested to create co-expression network from RNA-seq data. Certainly, a larger sample size can produce networks with a higher functional connectivity.
4. Pearson's correlation coefficient, Spearman's rank correlation coefficient, mutual information [34], and Euclidean distance are the four mostly common co-expression measures that are widely used in defining gene co-expression. Comparing with Pearson's correlation, etc., mutual information could also capture the nonlinear association [35].
5. Butte and Kohane used random permutations for the expression data to select a cutoff representing significant interactions [36]. Zhang and Horvath used soft thresholds instead of hard thresholds, to build weighted gene networks [37].
6. The clustering method should be selected carefully as it can greatly change the outcome and subsequently the biological meaning of this analysis. Several clustering methods are available, for example, hierarchical clustering and k-means clustering [38].

7. The default value of minModuleSize is 25, but here 5 is recommended.
8. The first PC accounts for the largest variance of the gene expression for the genes within the module and thus can describe the major expression pattern. This method is reasonable when the major variation in the data is caused by a treatment or condition. However, in practice, genes in the same module are not necessarily in the biological process due to different locations of gene products in cells, and mathematically, module-discovery methods may introduce large variance in the clustering process. The correlation approach based on single average patterns may fail to identify modules associated with the trait.
9. Linear model works with few modules and large samples. If the number of all eigengenes is large, then feature selection is necessary before the fitting step.
10. Wrong interpretation of low p -values may result in incorrect conclusion about the role of all genes enriched in a biological process [17, 18, 39].

Acknowledgments

The work is supported by funding under the National Science Foundation (Awards OIA-1557417 to CZ, DBI-1564621 to CZ HW and QZ, and OIA-1736192 to CZ HW and QZ). This work was completed utilizing the Holland Computing Center of the University of Nebraska.

References

1. Mao L, Van Hemert JL, Dash S et al (2009) Arabidopsis gene co-expression network and its functional modules. *BMC Bioinform* 10:346. <https://doi.org/10.1186/1471-2105-10-346>
2. Movahedi S, Van Bel M, Heyndrickx KS et al (2012) Comparative co-expression analysis in plant biology. *Plant Cell Environ* 35(10):1787–1798. <https://doi.org/10.1111/j.1365-3040.2012.02517.x>
3. Ruprecht C, Persson S (2012) Co-expression of cell-wall related genes: new tools and insights. *Front Plant Sci* 3:83. <https://doi.org/10.3389/fpls.2012.00083>
4. Yang Y, Han L, Yuan Y et al (2014) Gene co-expression network analysis reveals common system-level properties of prognostic genes across cancer types. *Nat Commun* 5:3231. <https://doi.org/10.1038/ncomms4231>
5. Tan M, Cheng D, Yang Y et al (2017) Co-expression network analysis of the transcriptomes of rice roots exposed to various cadmium stresses reveals universal cadmium-responsive genes. *BMC Plant Biol* 17(1):194. <https://doi.org/10.1186/s12870-017-1143-y>
6. Kadarmideen HN, Watson-Haigh NS (2012) Building gene co-expression networks using transcriptomics data for systems biology investigations: comparison of methods using microarray data. *Bioinformation* 8(18):855–861. <https://doi.org/10.6026/97320630008855>
7. Villa-Vialaneix N, Liaubet L, Laurent T et al (2013) The structure of a gene co-expression network reveals biological functions

- underlying eQTLs. PLoS One 8(4):e60045. <https://doi.org/10.1371/journal.pone.0060045>
8. Ficklin SP, Luo F, Feltus FA (2010) The association of multiple interacting genes with specific phenotypes in rice using gene coexpression networks. Plant Physiol 154(1): 13–24. <https://doi.org/10.1104/pp.110.159459>
 9. Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. BMC Bioinform 9:559. <https://doi.org/10.1186/1471-2105-9-559>
 10. Schafer J, Strimmer K (2005) An empirical Bayes approach to inferring large-scale gene association networks. Bioinformatics 21(6): 754–764. <https://doi.org/10.1093/bioinformatics/bti062>
 11. Gao C, McDowell IC, Zhao S et al (2016) Context specific and differential gene co-expression networks via bayesian biclustering. PLoS Comput Biol 12(7):e1004791. <https://doi.org/10.1371/journal.pcbi.1004791>
 12. Wang HQ, Tsai CJ (2013) CorSig: a general framework for estimating statistical significance of correlation and its application to gene co-expression analysis. PLoS One 8(10): e77429. <https://doi.org/10.1371/journal.pone.0077429>
 13. Watson-Haigh NS, Kadarmideen HN, Reverter A (2010) PCIT: an R package for weighted gene co-expression networks based on partial correlation and information theory approaches. Bioinformatics 26(3):411–413. <https://doi.org/10.1093/bioinformatics/btp674>
 14. Liang M, Zhang F, Jin G et al (2015) FastGCN: a GPU accelerated tool for fast gene co-expression networks. PLoS One 10(1):e0116776. <https://doi.org/10.1371/journal.pone.0116776>
 15. Wang Z, Fang H, Tang NL et al (2017) VCNet: vector-based gene co-expression network construction and its application to RNA-seq data. Bioinformatics 33(14): 2173–2181. <https://doi.org/10.1093/bioinformatics/btx131>
 16. Virilouvet L, Avenson TJ, Du Q et al (2018) Dehydration stress memory: gene networks linked to physiological responses during repeated stresses of *Zea mays*. Front Plant Sci 9:1058. <https://doi.org/10.3389/fpls.2018.01058>
 17. Di Salle P, Incerti G, Colantuono C et al (2017) Gene co-expression analyses: an overview from microarray collections in *Arabidopsis thaliana*. Brief Bioinform 18(2):215–225. <https://doi.org/10.1093/bib/bbw002>
 18. van Dam S, Vosa U, van der Graaf A et al (2018) Gene co-expression analysis for functional classification and gene-disease predictions. Brief Bioinform 19(4):575–592. <https://doi.org/10.1093/bib/bbw139>
 19. Borate BR, Chesler EJ, Langston MA et al (2009) Comparison of threshold selection methods for microarray gene co-expression matrices. BMC Res Notes 2:240. <https://doi.org/10.1186/1756-0500-2-240>
 20. Lee HK, Hsu AK, Sajdak J et al (2004) Coexpression analysis of human genes across many microarray data sets. Genome Res 14(6): 1085–1094. <https://doi.org/10.1101/gr.1910904>
 21. Stuart JM, Segal E, Koller D et al (2003) A gene-coexpression network for global discovery of conserved genetic modules. Science 302(5643):249–255. <https://doi.org/10.1126/science.1087447>
 22. Boyle EI, Weng S, Gollub J et al (2004) GO::TermFinder--open source software for accessing gene ontology information and finding significantly enriched gene ontology terms associated with a list of genes. Bioinformatics 20(18):3710–3715. <https://doi.org/10.1093/bioinformatics/bth456>
 23. Cline MS, Smoot M, Cerami E et al (2007) Integration of biological networks and gene expression data using Cytoscape. Nat Protoc 2(10):2366–2382. <https://doi.org/10.1038/nprot.2007.324>
 24. Obayashi T, Kinoshita K, Nakai K et al (2007) ATTED-II: a database of co-expressed genes and cis elements for identifying co-regulated gene groups in Arabidopsis. Nucleic Acids Res 35:D863–D869. <https://doi.org/10.1093/nar/gkl783>
 25. Ogata Y, Suzuki H, Sakurai N et al (2010) CoP: a database for characterizing co-expressed gene modules with biological information in plants. Bioinformatics 26(9): 1267–1268. <https://doi.org/10.1093/bioinformatics/btq121>
 26. Yim WC, Yu Y, Song K et al (2013) PLANEX: the plant co-expression database. BMC Plant Biol 13:83. <https://doi.org/10.1186/1471-2229-13-83>
 27. Netotea S, Sundell D, Street NR et al (2014) ComPIEx: conservation and divergence of co-expression networks in *A. thaliana*, *Populus* and *O. sativa*. BMC Genomics 15:106. <https://doi.org/10.1186/1471-2164-15-106>

28. Ma S, Bohnert HJ, Dinesh-Kumar SP (2015) AtGGM2014, an Arabidopsis gene co-expression network for functional studies. *Sci China Life Sci* 58(3):276–286. <https://doi.org/10.1007/s11427-015-4803-x>
29. Desai AP, Razeghin M, Meruvia-Pastor O et al (2017) GeNET: a web application to explore and share gene co-expression network analysis data. *PeerJ* 5:e3678. <https://doi.org/10.7717/peerj.3678>
30. Wong DC, Sweetman C, Drew DP et al (2013) VTCdb: a gene co-expression database for the crop species *Vitis vinifera* (grapevine). *BMC Genomics* 14:882. <https://doi.org/10.1186/1471-2164-14-882>
31. Hamada K, Hongo K, Suwabe K et al (2011) OryzaExpress: an integrated database of gene expression networks and omics annotations in rice. *Plant Cell Physiol* 52(2):220–229. <https://doi.org/10.1093/pcp/pcq195>
32. Lin H, Yu J, Pearce SP et al (2017) RiceAntherNet: a gene co-expression network for identifying anther and pollen development genes. *Plant J* 92(6):1076–1091. <https://doi.org/10.1111/tpj.13744>
33. Ballouz S, Verleyen W, Gillis J (2015) Guidance for RNA-seq co-expression network construction and analysis: safety in numbers. *Bioinformatics* 31(13):2123–2130. <https://doi.org/10.1093/bioinformatics/btv118>
34. Daub CO, Steuer R, Selbig J et al (2004) Estimating mutual information using B-spline functions—an improved similarity measure for analysing gene expression data. *BMC Bioinform* 5:118. <https://doi.org/10.1186/1471-2105-5-118>
35. Song L, Langfelder P, Horvath S (2012) Comparison of co-expression measures: mutual information, correlation, and model based indices. *BMC Bioinform* 13:328. <https://doi.org/10.1186/1471-2105-13-328>
36. Butte AJ, Kohane IS (2000) Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. *Pac Symp Biocomput*: 418–429
37. Zhang B, Horvath S (2005) A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* 4:Article17. <https://doi.org/10.2202/1544-6115.1128>
38. Wang Z, San Lucas FA, Qiu P et al (2014) Improving the sensitivity of sample clustering by leveraging gene co-expression networks in variable selection. *BMC Bioinform* 15:153. <https://doi.org/10.1186/1471-2105-15-153>
39. Serin EA, Nijveen H, Hillhorst HW et al (2016) Learning from co-expression networks: possibilities and challenges. *Front Plant Sci* 7:444. <https://doi.org/10.3389/fpls.2016.00444>



Chapter 21

Statistical Methods for the Quantitative Genetic Analysis of High-Throughput Phenotyping Data

Gota Morota, Diego Jarquin, Malachy T. Campbell, and Hiroyoshi Iwata

Abstract

The advent of plant phenomics, coupled with the wealth of genotypic data generated by next-generation sequencing technologies, provides exciting new resources for investigations into and improvement of complex traits. However, these new technologies also bring new challenges in quantitative genetics, namely, a need for the development of robust frameworks that can accommodate these high-dimensional data. In this chapter, we describe methods for the statistical analysis of high-throughput phenotyping (HTP) data with the goal of enhancing the prediction accuracy of genomic selection (GS). Following the Introduction in Sec. 1, Sec. 2 discusses field-based HTP, including the use of unoccupied aerial vehicles and light detection and ranging, as well as how we can achieve increased genetic gain by utilizing image data derived from HTP. Section 3 considers extending commonly used GS models to integrate HTP data as covariates associated with the principal trait response, such as yield. Particular focus is placed on single-trait, multi-trait, and genotype by environment interaction models. One unique aspect of HTP data is that phenomics platforms often produce large-scale data with high spatial and temporal resolution for capturing dynamic growth, development, and stress responses. Section 4 discusses the utility of a random regression model for performing longitudinal modeling. The chapter concludes with a discussion of some standing issues.

Key words Genetic gain, High-throughput phenotyping, Image data, Longitudinal trait, Quantitative genetics

1 Introduction

The predicted rise in global temperatures, increased variability of precipitation events, and increased competition for freshwater resources and arable land threaten to place unique constraints on global agriculture. Plant breeders in the twenty-first century will need to develop cultivars that are both high-yielding and resilient to climate change. The evaluation and development of breeding material requires a multifaceted approach, necessitating the consideration of multiple complex, and often interdependent traits. Successful germplasm development is not only dependent on the increase in the performance of breeding material that is achieved each cycle but also the amount of time before a cultivar is released

to the end-users. Moreover, the development of elite cultivars tolerant to abiotic stresses requires careful consideration of a suite of morphological and physiological traits that facilitate adaptation (e.g., plasticity and stability) to a range of environmental conditions. Thus, genetic improvement in this respect is a highly demanding process that requires extensive phenotypic evaluation in multiple environments. Advancements in sequencing have led to new genomic tools and have opened new avenues of research that aid breeders in their selection procedure. For instance, next-generation sequencing techniques such as genotyping-by-sequencing [1] have significantly increased the number of markers discovered and the number of individuals that can be sequenced, providing a cost-efficient tool for breeders to obtain the genotypic profiles of individuals.

In parallel with next-generation sequencing advancements, new statistical methods have been developed to enable utilization of the vast amount of available genomic information for selection purposes. This is known as genomic selection (GS), and its fundamental concept was first introduced by Meuwissen et al. [2]. GS predicts the performance of unobserved individuals based on the linkage disequilibrium between markers and causal loci and the genomic relatedness between observed and unobserved individuals. It has been shown that GS can increase genetic gain by reducing the number of cycles and the number of progeny that need to be phenotypically tested, thus reducing the cost of a breeding pipeline. Since Meuwissen et al. [2], there have been improvements in the prediction accuracy of GS through the incorporation of pedigree information [3–6], environmental covariates, and genotype by environment interactions [7–10].

One of the main advantages of GS over phenotypic selection is that phenotypic information is not required for the validation set. However, the acquisition of accurate phenotypic information is still a crucial component for the training or calibration set in the model building process. In other words, the phenotypic information of selection candidates is not used directly for selection, but the predictive ability of the models is negatively affected by the absence of accurate phenotypes. Obtaining precise phenotypic values is not trivial, but it is a critical part of genome-enabled breeding [11–13].

In recent years, high-throughput phenotyping (HTP) has become an emerging technology that can assist breeders in improving selection procedures and developing commercial cultivars more rapidly and efficiently [14]. In particular, image-based plant phenotyping enables frequent, non-destructive evaluation of multiple traits for a large number of plants with high precision. Image-based phenotyping offers several advantages, including being generally non-destructive, requiring low or no physical human labor input, being cost effective, and the ability to measure multiple traits at the

same time in different locations and at different developmental growth stages [11, 15].

There are a wide range of HTP platforms that have been developed for the purpose of providing dense phenotypic information [16]. Remote sensing and robotic systems developed in greenhouses and growth chambers have a high initial cost but can be fully automated. Alternatively, HTP data can also be collected in the field, as described later. In general, these field-based systems are associated with a high initial cost and also require a well-trained operator for collecting high-quality data [15]. Reynolds et al. [15] characterized and compared the available platforms in terms of associated costs and purposes. Despite such costs, a growing number of breeding programs are utilizing HTP platforms to better understand the genetics of quantitative traits and leverage these high-dimensional data to enhance selection. Specifically, HTP can be used to both generate dependent phenotypic variables for the training set in prediction models and provide additional information on genetic predictor variables in GS models, thereby improving prediction accuracies for conventional breeding targets, such as yield. Thus, this type of data can serve two main purposes: (1) as a primary trait response (e.g., plant height, canopy coverage, and number of leaves), and (2) as a covariate associated with the target trait response (e.g., yield). We discuss these points further in the following sections.

2 Field-Based High-Throughput Phenotyping Using UAV

In this section, we show that HTP accelerates plant breeding by improving the response to selection [17]. HTP methods allow us to measure plants efficiently and accurately via automatic or semi-automatic analysis of data collected by cameras and sensors [18]. Methods for measuring plants cultivated in a field are collectively known as field-based HTP. Field-based HTP enables the measurement of a large number of plots in an experimental field using cameras and sensors mounted on different platforms, such as unoccupied aerial vehicles (UAV) [19], carts [20], tractors [21], and gantry cranes. Field-based HTP not only improves the efficiency and accuracy of phenotyping of plants in a field but also makes it possible to evaluate traits that are difficult to measure with conventional phenotyping methods. In particular, the UAV is one of the most cost-effective and easy-to-use platforms [22, 23]. Although the type of camera or sensor that can be mounted on a UAV is restricted by the payload capacity of the UAV, light-weight and small cameras and sensors have been developed, and their precision and cost-efficiency have rapidly improved in recent years. The UAV is commonly equipped with digital cameras, multispectral cameras, and thermal infrared imagers in

field-based HTP [24, 25]. In contrast, hyperspectral cameras and LiDAR (light detection and ranging) are currently not commonly mounted on a UAV but are mainly ground-based HTP platforms [26–30] because of their weight, size, and cost. The commercialization of the UAV and related equipment is progressing in various fields, and various measurement devices will be available for HTP in the near future.

Plant characteristics that can be measured using UAV are roughly divided into three types of traits: (1) geometric traits, (2) spectral traits, and (3) physiological traits. For geometric traits, plant height, canopy cover, and canopy volume are measured mainly with RGB cameras or multispectral cameras [26, 28–36]. To measure these traits, a method called Structure from Motion (SfM) is used to estimate the three-dimensional (3D) structures of plants or plant canopies from a sequence of images acquired by a UAV. The structure is obtained using a set of data points, called a point cloud, in a 3D space. The 3D coordinate information of a point cloud is converted into a digital surface model (DSM) and an orthomosaic image. DSM is used for measuring plant 3D structural traits, such as plant height and canopy volume, while orthomosaic images are used for traits evaluated from above the ground, such as canopy cover. Lodging of plants can also be measured by DSM analysis [37]. The numbers and locations of flowers, blooms, and heads are also measured as geometric traits [38, 39]. In these studies, image-based machine learning has been used for the detection of target objects (i.e., flowers, blooms, or heads) from images acquired by UAV. Guo et al. [38] employed a two-step machine learning method for the detection of sorghum heads and attained high accuracy on various genotypes with different head morphologies and at different growth stages. Xu et al. [39] used a convolutional neural network to detect cotton blooms and estimated the 3D coordinates of the blooms using a dense point cloud constructed by SfM. These two studies demonstrated the potential of the combinatory use of image-based machine learning and HTP. Moreover, these studies suggest that simple but labor-intensive measurements, such as the monitoring of flowering and heading, can be performed on a much larger scale with HTP and image-based machine learning than with conventional methods.

For spectral traits, vegetation indices (VI) calculated from multispectral images acquired by UAV are used for evaluating vegetation properties, such as plant structure, biochemistry, and plant physiological and stress status [31, 33, 34, 40–48]. A large number of VIs have been proposed and have been used in ground-based platforms, aircraft, and satellite remote sensing. The fine spatial resolution of a UAV enables the removal of soil and shadow pixels from images and can improve the estimation of vegetative properties. Jay et al. [47] used 6-band multispectral cameras to evaluate

the structural and biochemical plant traits of green fraction, green area index, leaf chlorophyll content, and canopy chlorophyll and nitrogen contents, showing that the fine spatial resolution of the UAV always improved the estimation accuracy of these traits. Although multispectral images allow us to estimate various VIs better than RGB images, multispectral cameras are usually more expensive and have lower resolution than RGB cameras. To resolve this issue, Khan et al. [49] proposed a method for model-based estimation of VIs using RGB images. In this method, mean VI values were computed from the near infrared and red channels of corresponding plots, and then a deep neural network was trained with the RGB images as the input source and the VI values as the target output. A similar approach can be applied to the estimation of hyperspectral VIs from multispectral or RGB images and will be useful because hyperspectral cameras are usually much more expensive than multispectral cameras.

As for physiological traits, traits such as leaf chlorophyll content, protein content, biomass, crop vigor, nutrition status, and water status are measured by various methods including 3D construction and spectral VIs. A method that is specific to physiological traits is thermal infrared imaging, which enables the measurement of canopy temperature and can be used as a tool to indirectly evaluate the transpiration rate of a plant. Tattaris et al. [24] used a thermal infrared camera and a multispectral camera coupled with UAV to measure canopy temperature and the VI of wheat and found that data acquired by UAV generally exhibited stronger correlations with yield and biomass than data obtained from ground-based phenotyping. Ludovisi et al. [50] applied thermal infrared imaging to measure the canopy temperature of black poplar using UAV and found that the canopy temperature showed a good correlation with ground-truth stomatal conductance. Although the canopy temperature is an important indicator of stress status, it is extremely sensitive to small environmental changes, making it difficult to assess through slow ground-based methods [37]. HTP using UAV provides a good solution for this problem.

2.1 Application of HTP in Breeding Populations

When selecting breeding populations using HTP, two relatively simple methods are considered: (1) indirect selection and (2) index selection. Another method, selection based on prediction with HTP and genomic information, will be described later. When genetic correlations exist, selection for one trait will cause corresponding changes in other traits that are correlated [51]. This change in response due to genetic correlation is called a correlated response and may be caused by pleiotropy or linkage disequilibrium.

In the indirect selection method, a target trait, X , is selected indirectly by selection for trait Y , which has a genetic correlation with trait X and can be measured by HTP. It is possible to improve the selection efficiency of the target trait, the measurement of which would be costly, time consuming, or labor intensive, with traits readily measured by HTP. For example, Madec et al. [26] measured wheat plant height with HTP using LiDAR and UAV and found that it was highly correlated with the plant height measured at the ground level. They also demonstrated that heading date could be estimated based on a growth curve of plant height measured by LiDAR. Kyratzis et al. [44] evaluated the potential use of VIs acquired by UAV for durum wheat phenotyping and found that one index was significantly correlated with grain yield.

In the index selection method, a target trait, X , is selected based on an index calculated from phenotypes of a set of m traits, Y_s , related to the target trait. The simplest index is a linear combination expressed as

$$I = \sum_{j=1}^m b_j y_j$$

where b_j and y_j are the weight and phenotypic value of trait Y_j , respectively. If we consider b_j and y_j as the effect and state of marker j in a set of genome-wide markers, index selection becomes GS. The weight, which represents the relative importance of each trait, can be determined by multivariate regression. For example, Kefauver et al. [25] built a model regressing the grain yield on VIs acquired by HTP using UAV with stepwise regression and found that the regression model explained 77.8% of the grain yield variation. Yu et al. [27] performed hyperspectral imaging of a wheat canopy and used the resulting data to detect *Septoria tritici* blotch disease and to quantify the severity of infection. They used partial least squares regression to build a prediction model for severity and found that the accuracy of prediction (correlation between observed and predicted values) was 0.38–0.60 for three disease metrics. Non-linear relationships between trait X and a set of traits Y_s can also be modeled in a selection index. Various types of models, including known and ad hoc machine learning models, can be used for building an index. Thorp et al. [52] proposed a method for deriving daily evapotranspiration based on a daily soil water balance model named FAO-56 [53], which was derived from an index acquired by HTP using UAV, to evaluate and improve the crop water use efficiency of cotton varieties. Collectively, indirect or index selection based on traits measured by HTP has strong potential to streamline the selection of important agronomic traits, such as plant height, heading date, grain yield, and disease resistance.

2.2 Genetic Gain in HTP-Based Selection

HTP-based selection and GS can accelerate plant breeding by improving the efficiency of selection. Response to selection is an index for evaluating the efficiency of selection [54]. The response to selection R is defined as the difference between the mean phenotypic values ($\bar{y}_o$) of progeny generated from the selected parents and the mean phenotypic values ($\bar{y}_p$) of the parental population before selection.

$$R = \bar{y}_o - \bar{y}_p.$$

If we denote the heritability of a trait targeted in the selection as h^2 and define the selection differential as the product of the phenotypic standard deviation σ_p and selection intensity i in the parent population,

$$R = ih^2\sigma_p.$$

This is an important formula in breeding known as the “breeder’s equation.” If a breeder knows the heritability of the target trait h^2 and the standard deviation of the phenotype σ_p in the parent population, it is possible to calculate the expected response to selection R under intensity i . Using the definition of heritability, $h^2 = \sigma_g^2/\sigma_p^2$, we can rewrite the formula as

$$R = ih\sigma_g,$$

where σ_g is the square root of the genetic variance in the parent population.

Now we consider the case in which we select trait X indirectly by selecting for trait Y, measured with HTP. In this case, the response to selection of the indirect selection of trait X with trait Y is

$$R_{XY} = i_Y h_Y r_{XY} \sigma_{gX},$$

where i_Y is the selection intensity of trait Y, h_Y is the square root of the heritability of trait Y, r_{XY} is the genetic correlation between trait X and trait Y, and σ_{gX} is the square root of the genetic variance of trait X in the parent population. To improve the efficiency of selection with HTP, this value should be larger than the response to selection of direct selection of trait X, i.e., $R_X = i_X h_X \sigma_{gX}$. That is, the condition for improving the selection efficiency with HTP is

$$i_Y h_Y r_{XY} > i_X h_X.$$

When the selection intensities of the two traits are the same ($i_Y = i_X$), the following two conditions should be satisfied: (1) trait Y measured by HTP has a higher heritability than trait X, and (2) the genetic correlation between trait X and trait Y is high. With HTP, however, it is often possible to evaluate a large number of genotypes (strains or individuals) as compared with direct selection of trait X using a conventional phenotyping method.

Therefore, the selection intensity of trait Y can be increased compared to the selection intensity of trait X. If $i_Y > i_X$, even when the heritability of trait Y is not larger than that of trait X, it may be possible to perform indirect selection on trait X with higher efficiency than that of direct selection.

Index selection with HTP and GS both involve indirect selection of trait X based on the index I, which is calculated based on traits measured by HTP or genome-wide marker genotypes. The response to selection is represented as

$$R_{XI} = i_I r_{XI} \sigma_{gX},$$

where i_I is the selection intensity of the index I and r_{XI} is the accuracy of selection of trait X based on index I. The condition that the response to index selection is greater than the response to direct selection of trait X is

$$i_I r_{XI} > i_X h_X.$$

When the selection intensities of index I and trait X are the same ($i_I = i_X$) and the accuracy r_{XI} of selection of trait X based on index I exceeds the square root of the heritability of trait X, h_X , the efficiency of selection by index selection exceeds the efficiency of direct selection of trait X. As in the case of indirect selection using trait Y, if $i_I > i_X$, even when the accuracy r_{XI} of selection of trait X based on index I does not exceed the square root of the heritability of trait X, index selection has a higher efficiency than direct selection.

When we consider the efficiency of a breeding program, it is important to evaluate the genetic gain per unit time. Dividing the reaction to selection R by the time δ_X required for one cycle of selection, we obtain

$$\Delta G_X = \frac{i h_X \sigma_{gX}}{\delta_X},$$

where ΔG is the genetic gain per time. The genetic gain of indirect selection of trait X with trait Y is

$$\Delta G_{XY} = \frac{i h_Y r_{XY} \sigma_{gX}}{\delta_Y},$$

and the genetic gain of index selection of trait X with index I is

$$\Delta G_{XI} = \frac{i_I r_{XI} \sigma_{gX}}{\delta_I}.$$

Here, δ_Y and δ_I are the times required for one cycle of indirect and index selection, respectively. The time required for one cycle of selection can be shorter for trait Y and index I than for trait X. For example, the yield and quality of a grain crop are usually evaluated with multiple plants on a plot-by-plot basis. However, in indirect and index selection, it may be possible to perform selection

on a single plant basis in earlier generations, such as second-generation hybrids (F_2). In such a case, δ_Y (or δ_I) $< \delta_X$, and even when the response to selection R_{XY} or R_{XI} is smaller than the response to selection R_X , the genetic gain per unit time becomes greater under indirect and index selection than under direct selection.

As described above, the efficiency of selection can be improved by taking advantage of HTP, especially in terms of improvements in selection intensity and the time required for one cycle of selection. Field-based HTP is useful for increasing selection intensity because of its scalability, while HTP in the greenhouse is good for reducing the time required for one cycle of selection because it is often performed on a single-plant basis and year-round. In the application of HTP in plant breeding, the factors described earlier should be taken into account to optimize selection methods for target traits.

2.3 Use of HTP for GWAS and GS

Although HTP alone is expected to improve the response to selection, response to selection can be further improved by using HTP in combination with genome-wide association studies (GWAS) and GS. HTP with UAV is particularly suited for this purpose, as it can measure a large number of small- to medium-sized plots in which plants are cultivated. HTP with UAV has been applied to the evaluation of a large number of genotypes (germplasm accessions, varieties, and breeding lines) in many species, including wheat [26, 40, 42, 55, 56], maize [31, 33], sorghum [32, 36, 38], and black poplar [50]. Condorelli et al. [42] performed GWAS with 248 elite durum wheat lines to compare the results obtained with two UAVs and a ground-based method to measure a VI (Normalized Difference Vegetation Index, NDVI). More associations were detected by HTP using UAV than with the ground-based method, suggesting an improved ability of HTP using UAV over the ground-based method. Spindel et al. [36] undertook GWAS with 648 diverse sorghum lines for 460 combinations of traits, treatments, time points, and locations. Four traits related to biomass, plant height, and leaf area were measured by HTP using UAV. In total, 213 high-quality, replicated, and conserved associations were detected in genomic intervals, including many strong candidate genes. Watanabe et al. [32] measured the height of 115 sorghum germplasm accessions with HTP using UAV and evaluated the potential of HTP to provide phenotypic training data in a GS model. Although phenotypic correlation was not high, GS of plant height as measured by HTP using UAV was highly correlated with those measured manually. These results suggest the considerable potential of HTP using UAV for genomic-assisted breeding through GWAS and GS.

To successfully combine HTP with GWAS or GS, a novel viewpoint different from the analysis of conventional phenotypic data is necessary. Since HTP enables non-destructive and frequent measurements for large-scale field tests, a target trait can be measured as high-density time series data and as high-density data with coordinate information. Thus, spatial-temporal continuity and change can be taken into account in GWAS and GS models. For instance, Elias et al. [57] fitted a model with a spatial kernel as well as a kernel-based genomic relationship matrix to cassava agronomic trait data to account for the spatial heterogeneity in the field and showed that the prediction accuracy increased after accounting for the spatial variation. Moreover, multiple sensors are commonly employed in HTP, each of which can acquire high-dimensional data (e.g., hyperspectral images). Thus, for GWAS and GS using phenotypic data collected by HTP, it is necessary to consider the high dimensionality of the data and the large number of data points. Spindel et al. [36] conducted a GWAS on a number of features collected with HTP using UAV and constructed a method and pipeline to fuse and organize numerous GWAS results. Phenotypic data measured by HTP can also be used in the prediction of genotypic values of a target trait by leveraging genetic correlations between the target trait and traits measured by HTP. Rutkoski et al. [56] proposed a method for predicting a target trait with correlated HTP traits, as described in the next section.

3 Integration of HTP Data into GS

3.1 Single-Trait Analysis

Recently, there have been several studies that have integrated genomic data and HTP data for prediction purposes in several crops using different modeling techniques [13, 56, 58–63]. The integration of genomic and HTP data provides opportunities to improve existing GS models, thus enabling breeders to select their material more accurately and increase genetic gain. We summarize some key methods developed for integrating high-throughput genomic and HTP information for the purpose of increasing the accuracy of prediction by extending the standard GS models.

We can include secondary image traits in a quantitative genetics model using two model parameterizations. The first model explains the i th phenotypic observation as the sum of an intercept μ common to all observations, a linear combination of p markers x_{ij} and their corresponding marker effects b_j , a linear combination of Q secondary traits s_{iq} and their corresponding effects a_q , and residual ε_i as follows:

$$y_i = \mu + \sum_{j=1}^p x_{ij}b_j + \sum_{q=1}^Q s_{iq}a_q + \varepsilon_i.$$

The second model parameterization is based on covariance structures and can be obtained from the previous model by assuming that the effects of marker b_j and secondary traits a_q are independent and identically distributed draws from normal densities of the form $b_j \sim N(0, \sigma_b^2)$ and $a_q \sim N(0, \sigma_a^2)$. Then, $g_i = \sum_{j=1}^p x_{ij} b_j$ and $w_i = \sum_{q=1}^Q s_{iq} a_q$ are genetic and environmental values of the i th genotype using information from genomics and secondary traits. From properties of the multivariate normal density, the vectors of marker and secondary trait effects are also normally distributed, such as $\mathbf{g} = \{g_i\} \sim N(\mathbf{0}, \mathbf{G}\sigma_g^2)$ and $\mathbf{w} = \{w_i\} \sim N(\mathbf{0}, \mathbf{C}\sigma_A^2)$, where $\mathbf{G} = \mathbf{X}\mathbf{X}'/p$ is a covariance matrix whose entries describe genomic similarities between pairs of genotypes; $\mathbf{X}$ is the matrix of molecular markers of order $n \times p$; $\sigma_g^2 = p \times \sigma_b^2$; $\mathbf{C} = \mathbf{S}\mathbf{S}'/Q$ is a covariance matrix whose entries describe phenotypic similarities based on image secondary traits data for each pair of observations; $\mathbf{S}$ is the matrix of secondary traits of order $n \times Q$; and $\sigma_A^2 = Q \times \sigma_a^2$. This parameterization assumes that all of the secondary traits equally contribute to explain the phenotypic variations of the traits of interest. One of the advantages of using this second parameterization is that it is possible to evaluate the contribution of the genomic and HTP components for explaining phenotypic variability by comparing the estimated variance components associated with each of these terms.

The majority of models developed focus on predicting a single trait, namely, grain yield. HTP can measure traits that are shown to be highly correlated with grain yield, such as the spectral reflectance of the canopy and canopy temperature [64]. A VI is used to summarize the spectral reflectance of the canopy scores [61]. However, because the VI is calculated using only a subset of the available wavelengths, it does not take advantage of all of the HTP data. There are several approaches for incorporating all of the HTP wavelengths and the plot-level VI measurements into GS models. Rutkoski et al. [56] showed that the integration of VI and canopy temperature into a genomic best linear unbiased prediction (GBLUP) model could increase the prediction accuracy by 70% compared to that of a univariate baseline model in wheat data. Aguete et al. [65] showed that using bands as predictors increased prediction accuracy over that of VI. They used ordinary least squares, partial least squares, and a Bayesian shrinkage model to incorporate wavelengths into a GS model in maize. A similar observation was made by Montesinos et al. [66], who compared prediction model performance when all of the wavelengths were incorporated with that of a subset of the wavelengths in wheat. They concluded that using all of the wavelengths resulted in higher prediction accuracy.

3.2 *Multi-Trait Analysis*

Sun et al. [59] predicted grain yield in a two-step procedure in wheat data. First, they collected data on canopy temperature and VI as secondary traits (which are correlated with grain yield) and modeled the secondary traits using the genetic marker and environmental effects. They applied a mixed model for predicting grain yield without considering the secondary traits as covariates. However, they used the secondary traits to develop a multivariate model to predict grain yield, which is the primary trait. The secondary traits were measured in a longitudinal fashion, i.e., at several time points throughout the growing season. They implemented and compared the repeatability, multi-trait, and random regression (RR) models that can be used for modeling longitudinal data. In the second step of the GS, the results from the repeatability, multi-trait, and RR models were used as BLUP, and a univariate prediction model was compared to bivariate and multivariate models. Only grain yield was included, and the secondary traits were excluded in the univariate model. In one of the multivariate prediction models, the secondary traits were included both in the training and testing sets, and in the other multivariate prediction model the secondary traits were included only in the training set. The bivariate prediction model included grain yield and one of the secondary traits. Their results showed that the multivariate prediction model that incorporated the secondary traits in both the testing and training sets had an advantage over the other models in terms of prediction accuracy. However, it was not clear which of the first models (repeatability, multi-trait, or RR) performed the best because the results depended on the environmental conditions. Nonetheless, the results clearly demonstrated the advantage of using HTP data in GS applications.

Crain et al. [13] compared four models using wheat data: (1) a regular GS model, (2) a univariate model in which grain yield was the response and HTP data were predictors, (3) a model that was the combination of models 1 and 2, and (4) a multi-trait model that included grain yield, canopy temperature, and VI measurements. The results showed that the addition of HTP data increased the prediction accuracy. They found that the multi-trait model exhibited a 7% gain in terms of prediction accuracy, indicating that collecting multiple HTP measurements has the potential to increase genetic gain through the improvement of prediction models. Juliana et al. [62] applied multivariate prediction models to compare standard GS with a pedigree- and HTP-based prediction model. They discussed the situations in which each model can be useful and the importance of implementing the correct models in the correct stage of the breeding pipeline. The authors elaborated on the importance of the family structure and of the secondary HTP traits being highly correlated with the primary phenotypic trait, as these components are influencing factors in prediction performance.

3.3 Genotype by Environment Interaction

Although all of the studies described above considered approaches to integrate HTP into GS, they did not apply interaction effect models. However, there are multiple lines of evidence that GS models with interaction effects have the potential to outperform competing models with only additive effects [67–69]. Montesinos et al. [70] presented one of the first studies of HTP showing the impact of including the interaction between hyperspectral bands and environments (band $\times$ environment). These authors found that the model with the band $\times$ environment interaction outperformed all of the models without this interaction term. Jarquin et al. [63] used prediction models that incorporated line, environment, marker genotype, canopy coverage image information, and their interactions in soybean. They evaluated six main effects' models that included combinations of line, environment, marker genotype, and canopy coverage image information; seven models with two-way interactions among the components; and two models with a three-way interaction between environment, marker genotype, and the canopy coverage data. Under the GBLUP model, they modeled the interaction components as the Hadamard product [71] of the relationship matrices obtained from genetic marker and canopy coverage image information according to the reaction norm model [9]. The model performance was evaluated using three cross-validation (CV) schemes: CV2, CV1, and CV0. CV2 assumed an incomplete field trial, in which some lines are observed in some environments but not in others. CV1 was the case in which one predicts the performance of a new line in environments in which some other lines were evaluated. The goal of CV0 was to predict the performance of already tested lines in untested environments. When grain yield was the target trait, the advantage of including the canopy coverage measurements and the interactions among marker, environment, and canopy coverage measurement was clearly shown. The highest predictive abilities for CV2 and CV1 were delivered by the models that included a three-way interaction among marker genotype, canopy coverage image data, and environment, while for CV0, the model with interactions between marker genotype and environment, and between canopy coverage image information and environment produced the greatest accuracy. The study also evaluated the effectiveness of canopy coverage image data from early stages and compared it with the case in which the canopy coverage image data was collected throughout the growing season. The results indicated that the information collected in the early stages was sufficient for prediction and that the additional data collected in the later stages did not improve the prediction models significantly. The practical implication of this finding is important, as it shows that the same prediction accuracy can be achieved using fewer resources (time, measurements, and costs).

Krause et al. [61] used multi-kernel, multi-environment GBLUP models including genetic marker or pedigree, environmental, and hyperspectral band information for predicting grain yield in wheat. They found that when marker genotype or pedigree data are not available, the main effects model using the hyperspectral band data provided a similar accuracy of prediction compared to the main effects models including marker or pedigree information. Additionally, the model with interactions outperformed the main effects models. Their findings differed from those of Jarquin et al. [63] with regard to the effectiveness of including partial HTP data. They concluded that the prediction accuracy increased when the HTP data from later stages were included. However, this difference is expected, as the crop development for wheat is significantly different from that for soybean. Finally, Montesinos et al. [70] and Montesinos et al. [72] showed the advantages of performing functional analysis for reducing data dimensionality to extract a higher signal-to-noise ratio for each observed value. In addition, Montesinos et al. [70] showed that when the HTP collected over multiple time points are combined using functional analysis, a small increase in prediction accuracy can be achieved relative to that of models that use data from a single time point.

4 Utilizing Image-Derived Longitudinal Traits for Genetic Studies in Plants

The observable phenotype at a given time is the culmination of numerous biological processes that have occurred prior to observation. For example, consider a cereal such as wheat at maturity. The total above-ground biomass can be separated hierarchically into a number of distinct organs. The whole plant can be partitioned into main and auxiliary tillers, which can be further partitioned into leaf blades, leaf sheaths, and stems. This process can proceed further to lower organization levels, separating these organs into tissues and cellular components. At each level, the pattern timing of development is tightly controlled by complex genetic networks that, at the organ level, control the onset of primordial development and initiation of growth and, at the plant level, the transition from vegetative to reproductive development.

An additional layer of complexity is added to this when the effect of the environment on these processes is considered. The appearance of the plant at maturity is certainly a product of its genetic makeup; however, the processes mentioned above are all tightly linked to the environment. The total size of the plant at maturity is a product of the resources (e.g., light, nutrients, and water) that were available throughout its life cycle. Plants need light and carbon dioxide to produce sugars through photosynthesis. Nutrients are combined with these sugars to generate nucleotides, proteins, and metabolites. Limitations on any of these inputs will

slow or stunt growth. In addition to plant growth, the transition between developmental states is also linked directly to the environment. Several studies have shown that drought events can lead to earlier flowering and accelerated post-anthesis development (reviewed by Shavrukov et al. [73]). Therefore, the phenotype is not a static entity. The observable phenotype is the result of dynamic genetic processes, the changing external environment, and the dynamic interplay between the two.

For most genetic applications, plants are often phenotyped at only one or a few time points. These phenotypes are an incredibly informative summation of the processes that have occurred over the life cycle of the plant, and they have been used quite successfully to select for a variety of complex traits. While for many applications these single time point phenotypes may be sufficient, they fail to capture the dynamic processes that have led to the observable phenotype. In most genetic studies, phenotypic evaluation is the largest, most time-consuming activity. Typical genetic studies consist of a mapping population with hundreds to thousands of individuals that are grown in replicates. Thus, for these studies, phenotyping at one time point is often a huge commitment, while evaluation at multiple time points is often unfeasible.

In the last decade, the construction and accessibility of high-throughput phenotyping platforms have provided an attractive means for generating phenotypic data throughout the duration of a study in a non-destructive manner for a number of economically important crop species [14, 32, 37, 74]. These platforms have been successfully deployed in controlled environments to quantify growth and physiological processes in response to drought and salinity [75, 76]. Moreover, with the growing popularity of UAVs in the consumer market, a vast selection of hardware can be obtained at relatively low cost [32]. These can be outfitted with various sensors or cameras and deployed routinely in the field to capture trait development over the growing seasons. In crop species, these temporal phenotypes have been used as covariates in genomic prediction frameworks to improve prediction accuracy for end-point phenotypes, such as yield [13, 59, 63]. However, analysis of the longitudinal trait itself has been largely confined to genetic inference in crops species, while genomic prediction has been applied largely to perennial species [77–79]. In the following section, we describe several approaches for genomic prediction of the longitudinal phenotype itself.

4.1 Single Time Point Genetic Inference

A seemingly straightforward approach for assessing dynamic genetic effects underlying longitudinal traits is performing linkage or association analysis at each time point independently [80–83]. In one of the first applications of HTP for genetic studies in plants, Moore et al. [83] used an image-based platform to quantify root gravitropic responses in *Arabidopsis* biparental mapping

populations. The authors used a step-wise mapping approach at each time point to identify time-dependent quantitative trait loci (QTL) and used a post hoc approach to combine information on QTL detected across multiple time points. The post hoc approach effectively used two metrics to classify QTL into a persistent class, by averaging the LOD scores across time points, and transient QTL, by taking the maximum LOD across all time points. While this post hoc approach effectively combines statistics across time points and successfully classifies the temporal genetics of root gravitropism, the single time point mapping approach itself does not explicitly model the covariance across time points. Thus, the actual genetic inference step does not fully capture the phenotypic trajectories.

4.2 Functional Mapping

Several other approaches have been proposed that directly consider the trait trajectories for genetic analyses. With these approaches, the trait values across all time points can be modeled using parametric or non-parametric mathematical functions. These models describe the phenotypic trajectories using a few parameters (for a review of parametric models in the context of plant growth, see Paine et al. [84]). Once an appropriate model has been chosen, genetic inference or prediction can proceed using a single-step or two-step approach.

4.2.1 Single-Step Functional Mapping

In the single-step functional mapping approach, model fitting and genetic analyses are performed within a single statistical framework. In the plant community, the single-step approach for functional genetic inference/mapping was first proposed by Ma et al. [85] to map QTL for stem diameter in *Populus*. Since then, the functional mapping approach has been applied to longitudinal traits in other species, such as humans and mice, and has been extended into the mixed model framework used for GWAS [86–90]. The advantages of the single-step functional mapping approach are that it considers the full trait trajectories over time, yielding loci that influence the curve itself, and captures the covariance across time points, which should reduce residual variance and improve statistical power [88]. Essentially, at each locus, the single-step functional mapping approach models the mean trajectories for each genotype and tests whether the time-dependent genetic effects are non-zero.

There are two important considerations for the single-step functional mapping approach: (1) the choice of function to model the mean trajectories of each genotype, and (2) the appropriate residual covariance structure to account for the temporal nature of the data. The function to model the mean trajectories can be parametric or non-parametric and can be selected based on some prior knowledge of the phenotypic trajectories. For well-studied traits, such as growth, a number of parametric options exist, are biologically meaningful and can be easily applied to the

longitudinal data set [84]. In cases in which no prior knowledge exists about the phenotypic trajectories, a nonparametric function, such as orthogonal Legendre polynomials or B-spline functions, can be utilized. The nonparametric functions are described in greater detail below. A number of covariance structures can be used to account for the temporal relationships among observations. The choice will be dependent on the balance between statistical efficiency and robustness. In the most robust case, the unstructured covariance matrix, the variance and covariance at each time point are unique and estimated from the data. While this places no constraints on the variance–covariances, the number of parameters that must be estimated can be prohibitively large for most studies. In many cases, simpler structures may be nearly as robust while estimating far fewer parameters.

4.2.2 Two-Step Functional Mapping

In contrast to the single-step functional mapping approach, the two-step approach performs the model fitting and genetic analysis in two separate steps. First, the phenotypic trajectories are modeled for each individual, and the model parameters are used as derived traits for subsequent genetic analyses (e.g., GWAS, linkage analysis, or GS). This two-step approach has been successfully used to examine the genetic basis of rosette growth in *Arabidopsis* and for GWAS and GS of early vigor in rice [91, 92]. The advantages of this approach are that it is conceptually simple and easy to implement. Moreover, for most popular growth models, the parameters have biological meaning. For instance, growth can be modeled over the life cycle of the plant using a 3-parameter logistic function. Here, the inflection point can be calculated, which represents the transition from vegetative to reproductive growth. Thus, the researcher can select a specific attribute to study and select a specific model parameter that represents that attribute for analysis. Moreover, outside of genetic mapping, these parameters may provide biological insight into a plant's phenotypic development. For instance, Campbell et al. [92] targeted a specific model parameter that described a plant's growth rate and showed that the plant hormone gibberellic acid may influence natural variation for the rate-controlling parameter. However, the major disadvantage of this method is that information is lost between the functional modeling and genetic analysis steps. Since environmental factors are not included in the functional modeling step, the residuals likely contain important information regarding non-genetic components of the phenotypic variance for the longitudinal phenotype.

4.3 Insights from Animal Breeding for Genomic Prediction Using Longitudinal Traits

While the use of longitudinal phenotypes is relatively new in plant science, animal breeders have targeted longitudinal traits for decades [93]. In animal breeding, breeders are often interested in the development of a trait across an animal's life. For instance, in dairy cattle, test-day milk yields are collected routinely. Moreover, other traits, such as feed intake, growth, and egg production [94–97],

have also been examined in a longitudinal framework. With the extensive use of these traits in animal breeding, numerous frameworks have been well developed to accommodate the time axis and have been used extensively for inference on genetic and environmental variance components, as well as pedigree and GS.

In the following subsection, we discuss several approaches that have been used for pedigree- or genomic-based prediction in animal breeding in a context that is applicable to plant breeding with HTP platforms. As mentioned above, a naive approach for GS using longitudinal data would be a univariate approach, in which a conventional mixed model is fitted at each time point. Here, we introduce the concept of longitudinal GS from a multivariate framework, as this is a relatively simple extension of the univariate approach, and extend these concepts to covariance functions and RR models that have been pioneered in animal breeding.

4.4 Multivariate Approaches for Longitudinal Genomic Prediction

To capture the covariance between time points, a logical progression from the univariate approach is to utilize a multivariate framework for longitudinal data. Thus, rather than considering the longitudinal trait as a consecutive series of measurements on the same trait, with the multivariate approach, we essentially ignore the order of the series and treat each time point as a separate trait. The multivariate framework allows each time point to have a unique variance and unique covariances between time points. The multivariate GS framework is well developed and has been widely utilized in both plant and animal systems. Moreover, the extension from the univariate approach is relatively straightforward.

Assume a simple case in which we are given three consecutive measurements for each individual. The model for each trait can be written as

$$y_1 = X_1 b_1 + Z_1 u_1 + \varepsilon_1 \quad (1)$$

$$y_2 = X_2 b_2 + Z_2 u_2 + \varepsilon_2 \quad (2)$$

$$y_3 = X_3 b_3 + Z_3 u_3 + \varepsilon_3 \quad (3)$$

where y_i is the vector of observations for trait i ; X_i and Z_i are the incidence matrices for fixed effects and random effects, respectively, for trait i ; u_i is the vector of random genetic effects for trait i ; and ε_i is the vector of residuals for trait i . Thus, the multivariate model is

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} X_1 & 0 & 0 \\ 0 & X_2 & 0 \\ 0 & 0 & X_3 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} + \begin{bmatrix} Z_1 & 0 & 0 \\ 0 & Z_2 & 0 \\ 0 & 0 & Z_3 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix} \quad (4)$$

Moreover, as mentioned above, we assume unique variances and covariances for each trait/time point.

$$\text{var} \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \mathbf{u}_3 \\ \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix} = \begin{bmatrix} \mathbf{G}\sigma_g^2_{11} & \mathbf{G}\sigma_g^2_{12} & \mathbf{G}\sigma_g^2_{13} & 0 & 0 & 0 \\ \mathbf{G}\sigma_g^2_{21} & \mathbf{G}\sigma_g^2_{22} & \mathbf{G}\sigma_g^2_{23} & 0 & 0 & 0 \\ \mathbf{G}\sigma_g^2_{31} & \mathbf{G}\sigma_g^2_{32} & \mathbf{G}\sigma_g^2_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{I}\sigma_\varepsilon^2_{11} & \mathbf{I}\sigma_\varepsilon^2_{12} & \mathbf{I}\sigma_\varepsilon^2_{13} \\ 0 & 0 & 0 & \mathbf{I}\sigma_\varepsilon^2_{21} & \mathbf{I}\sigma_\varepsilon^2_{22} & \mathbf{I}\sigma_\varepsilon^2_{23} \\ 0 & 0 & 0 & \mathbf{I}\sigma_\varepsilon^2_{31} & \mathbf{I}\sigma_\varepsilon^2_{32} & \mathbf{I}\sigma_\varepsilon^2_{33} \end{bmatrix} \quad (5)$$

Thus, for this simple case, we are capturing the full covariance across the three time points and leveraging this covariance to predict unique genetic values at each. However, notice the dimensions of the covariances σ_g^2 and σ_ε^2 . Here, we must solve for 12 parameters. If we have a very large population, this may not be an issue. However, if we consider a more realistic data set from HTP, it is likely that we will have many more time points. Thus, for t time points, we will need to estimate t variances and $t(t-1)/2$ covariances for both the genetic effects and residuals. For most HTP studies, this will create unnecessary computational demands. Moreover, additional challenges could be experienced if the parameter estimates are near the bounds, which may yield inaccurate estimates of variance components. Thus, when faced with larger longitudinal data sets ($t > 5$), the researcher should question whether it is necessary to estimate each covariance. In cases in which the measurements are taken at short intervals within a given developmental period, it is likely safe to assume that the genetic variances between adjacent time points will be very similar. Therefore, a much simpler model may still capture much of the covariance while estimating fewer parameters. This is discussed in detail below. For other cases in which fewer measurements are recorded over more widely spaced intervals, the previous assumption may not hold true, and the full, unstructured matrix used in the multi-trait framework may be a more accurate model.

4.5 Covariance Functions and Random Regression Models for Longitudinal Genetic Prediction

In the multi-trait framework, we treat the longitudinal phenotype, say growth, as a collection of independent traits; as a result, we are limited to making predictions at time points with records. However, in most longitudinal studies, we are interested in learning about the development of a continuous trait over time and do so by taking measurements at discrete time points. The time points or intervals themselves may be selected somewhat arbitrarily, and we seek to fill in information between time points. Thus, to capture the full trajectory of trait development, we can separate the trajectory into infinitely smaller intervals. Therefore, if we view the longitudinal trait as an “infinite-dimensional” trait, we can see that the

multivariate framework is inadequate, in that it does not directly consider the time axis and it does not allow us to make predictions at time points without observations.

Kirkpatrick et al. [98] initially proposed the use of covariance functions (CFs) for the analysis of “infinite-dimensional” traits. A CF is simply the infinite-dimensional equivalent of a covariance matrix for a given number of time points. Using this approach, the covariance between any two records measured at given time points can be obtained using only the time points and some coefficients. For an “infinite-dimensional” trait, there can be an infinite number of coefficients; however, in practice, the number of coefficients is dependent on the number of time points with records, with the maximum number of coefficients being $t(t+1)/2$.

Following the example described in the multi-trait section above, we provide a brief example of the CF approach. Assume we have a trait measured at three time points using the covariance matrix in Kirkpatrick et al. [98]. Using the multi-trait approach, we estimate the 3×3 additive genetic covariance matrix ($\hat{\Sigma}$) and estimate the variances and covariances at each of the three time points. The goal of the approach described by Kirkpatrick et al. [98] is to represent the additive generic covariance matrix ($\hat{\Sigma}$) as a continuous covariance function ($\mathcal{K}$) given data collected at discrete time points. Although a number of methods can be used to estimate $\mathcal{K}$ from $\hat{\Sigma}$, orthogonal polynomials are used most often due to the low correlations among the estimated coefficients [99].

Given a covariance function with a full rank fit (e.g., order of polynomials is equal to the number of time points, $k=t$), Kirkpatrick et al. [98] showed that the observed covariance matrix $\hat{\Sigma}$ can be expressed as $\hat{\Sigma} = \Phi \mathbf{K} \Phi'$, where $\mathbf{K}$ is a coefficient matrix associated with the CF, and Φ is a matrix of Legendre polynomials of order t by k , the order of Legendre polynomials (in this case $k=t$). Φ is defined by the Legendre polynomial functions via $\Phi = \mathbf{M} \mathbf{\Lambda}$. With Legendre polynomials, the time points are standardized so that they span an interval of -1 to 1, and here, $\mathbf{M}$ is a matrix of the polynomials of standardized time points. $\mathbf{\Lambda}$ is a matrix of coefficients of Legendre polynomials of order $k \times k$. The first two Legendre polynomials are $P_0(t)=1$ and $P_1(t)=t$, and the subsequent j th Legendre polynomials are given by $P_{j+1}(t) = \frac{1}{j+1}(2j+1)tP_j(t) - jP_{j-1}(t)$. These can be normalized to ϕ_j via $\phi_j = \frac{\sqrt{(2j+1)}}{2} P_j(t)$. Thus, the first three normalized Legendre polynomials will be $P_0(t)=0.707$, $P_1(t)=1.2247t$, and $P_2(t)=-0.7906+2.3717t^2$. Thus, $\mathbf{\Lambda}$ is

$$\mathbf{\Lambda} = \begin{bmatrix} 0.7071 & 0 & -0.7906 \\ 0 & 1.2247 & 0 \\ 0 & 0 & 2.3717 \end{bmatrix} \quad (6)$$

It is of particular importance to note that Φ is not dependent on the values nor the time points in the data set; only $\mathbf{K}$ is. Thus, given the 3×3 covariance matrix, $\hat{\Sigma}$, the covariance between any two time points, can be obtained using $\mathcal{K}(a_1, a_2) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \mathbf{K}_{ij} \phi_i(a_1) \phi_j(a_2)$, and the breeding value at any time point can be obtained using $t = \sum_{i=0}^{k-1} \phi_i(a_t) u_i$. Moreover, with a full rank fit, the covariance matrix obtained is equivalent to that obtained using the multivariate approach in the previous section.

In most cases, the full covariance matrix $\hat{\Sigma}$ is unknown; therefore, it must be estimated directly from the data. As shown by Meyer and Hill [100], this can be done by a reparameterization of the multivariate or “finite-dimensional” approach. However, in many studies, particularly those focused on the analysis of longitudinal milk production in dairy cattle, RR models (e.g., test-day models) are most commonly used. The RR approach proposed by Schaeffer [93] regresses the phenotypic trait trajectories directly on Φ to estimate $\mathbf{K}$. As demonstrated by Meyer and Hill [100], both the CF and RR approaches are equivalent. The general form of the RR model is

$$y_{tij} = FE_i + \sum_{k=0}^{nf} \phi_{jtk} \beta_k + \sum_{k=0}^{nr} \phi_{jtk} u_{jk} + \sum_{k=0}^{nr} \phi_{jtk} pe_{jk} + \varepsilon_{tij} \quad (7)$$

Here, FE_i is the fixed effect for the i th group; ϕ_{jtk} is the k th Legendre polynomial for individual j at time t ; β_k is the fixed regression coefficient for the k th Legendre polynomial, which represents the overall mean trait trajectory for the population or group; u_{jk} is the genetic value for the k th Legendre polynomial for the j th individual; and pe_{jk} is the permanent environmental effect for the k th Legendre polynomial for the j th individual. This permanent environmental effect is a stable, perpetual, non-genetic effect that influences an individual’s trait trajectory. It is assumed to be common to all repeated observations on the same individual. Thus, $\mathbf{e}$ can be considered temporary environmental effects. In matrix form, the RR model can be written as $\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{Qp} + \boldsymbol{\varepsilon}$.

In the examples above, we used a full-order polynomial to model the covariance across time points. As in the multivariate example, this requires estimation of a large number of parameters and in most cases is computationally unfeasible and could lead to convergence problems or inaccurate parameter estimates. In most cases, it is much more advantageous to fit a simpler model using a reduced-order polynomial ($k < t$). This effectively allows fewer

parameters to be estimated while still adequately describing $\hat{\Sigma}$. Generally speaking, the goodness of fit will increase as the number of function parameters describing the curve increases [101]. Campbell et al. [102] used RR models for rice shoot growth trajectories and demonstrated that the model could be used for longitudinal genomic prediction. Baba et al. [103] showed the utility of a multi-trait RR model for genomic prediction of daily water usage in rice through joint modeling with shoot biomass.

5 Conclusions

This chapter described statistical methods for analyzing large-scale HTP data in quantitative genetics. We contend that the integration of HTP data into quantitative genetics models triggers a great leap forward in plant breeding. In particular, we discussed (1) the genetic gain that can be achieved using HTP data, (2) the use of HTP data as predictive covariates in GS models, and (3) the modeling of temporal HTP data using RR models. In GS, it is known that the accuracy of genomic prediction, and thus the response to selection, decreases as the selection cycle advances [104, 105]. To maintain the response to selection, it is necessary to update the model on a regular basis [105–107]. In order to update the model, it is necessary to conduct a field test to measure phenotypes and to obtain genome-wide markers for many genotypes. At this step, phenotypic measurement for model updating may become a serious bottleneck of GS breeding. Thus, it is important to utilize HTP, which can evaluate many genotypes and possibly shorten the time required for selection.

High-throughput phenotyping and phenomics offer numerous opportunities to understand plant development, the genetics of quantitative traits such as yield, and their connection to the environment. The utilization of HTP data that are correlated with traits of interest can change how breeders select their material for advancement. Incorporating HTP data into prediction models has the potential to increase prediction accuracy, thus enabling plant breeders to select and discard more accurately. Although the reviewed studies considered different models, they concluded that regardless of the model configuration, the inclusion of HTP data increased the prediction performance when it was combined with different data types (marker genotype, pedigree, and environment). Additional gains can be expected when considering interactions with environmental factors.

The RR approach offers several advantages compared to the multivariate approach. As mentioned above, the RR approach allows environmental effects to be partitioned into permanent and temporary environmental effects. Moreover, the RR approach models the individual-specific deviations from the mean phenotypic

trajectories of the population. This allows the shape, amplitude, and intercept of the phenotypic trajectories to be unique for each individual and assumes that the genetic and permanent environmental effects are not constant throughout trait development. Thus, the RR model should more accurately reflect the biological processes that give rise to the phenotype. Furthermore, RR models offer a robust framework for fitting reduced-fit covariance functions. This offers a computational advantage over the multivariate approach in that it allows the model to converge more quickly. Moreover, by only estimating the parameters that are necessary to describe the data, sampling errors can be minimized. Finally, the RR approach provides a robust framework that allows the researcher to study how genetic variability changes over time and enables selection of individuals to alter phenotypic trajectories over time.

References

1. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS One* 6(5):e19379
2. Meuwissen T, Hayes B, Goddard M (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157(4):1819–1829
3. De Los Campos G, Naya H, Gianola D, Crossa J, Legarra A, Manfredi E, et al (2009) Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics* 182(1):375–385
4. Crossa J, Campos Gdl, Pérez P, Gianola D, Burgueño J, Araus JL, et al (2010) Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics* 186(2):713–724
5. Aguilar I, Misztal I, Johnson D, Legarra A, Tsuruta S, Lawlor T (2010) Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J Dairy Sci* 93(2):743–752
6. Christensen OF, Lund MS (2010) Genomic prediction when some animals are not genotyped. *Genet Sel Evol* 42(1):2
7. Burgueño J, de los Campos G, Weigel K, Crossa J (2012) Genomic prediction of breeding values when modeling genotype $\times$ environment interaction using pedigree and dense molecular markers. *Crop Sci* 52(2):707–719
8. Heslot N, Akdemir D, Sorrells ME, Jannink JL (2014) Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. *TheorAppl Genet* 127(2):463–480
9. Jarquín D, Crossa J, Lacaze X, Du Cheyron P, Daucourt J, Lorgeou J, et al (2014) A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theor Appl Genet* 127(3):595–607
10. Pérez-Rodríguez P, Crossa J, Rutkoski J, Poland J, Ravi S, Legarra A, et al (2017) Single-step genomic and pedigree genotype $\times$ environment interaction models for predicting wheat lines in international environments. *Plant Genome* 10(2)
11. White JW, Andrade-Sanchez P, Gore MA, Bronson KF, Coffelt TA, Conley MM, et al (2012) Field-based phenomics for plant genetics research. *Field Crops Res* 133:101–112
12. Cobb JN, DeClerck G, Greenberg A, Clark R, McCouch S (2013) Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype–phenotype relationships and its relevance to crop improvement. *Theor Appl Genetics* 126(4):867–887
13. Crain J, Mondal S, Rutkoski J, Singh RP, Poland J (2018) Combining high-throughput phenotyping and genomic information to increase prediction and selection accuracy in wheat breeding. *Plant Genome* 11(1):170043

14. Furbank RT, Tester M (2011) Phenomics—technologies to relieve the phenotyping bottleneck. *Trends Plant Sci* 16(12):635–644
15. Reynolds D, Baret F, Welcker C, Bostrom A, Ball J, Cellini F, et al (2019) What is cost-efficient phenotyping? Optimizing costs for different scenarios. *Plant Sci* 282:14–22
16. Fiorani F, Schurr U (2013) Future scenarios for plant phenotyping. *Ann Rev Plant Biol* 64: 267–291
17. Araus JL, Kefauver SC, Zaman-Allah M, Olsen MS, Cairns JE (2018) Translating high-throughput phenotyping into genetic gain. *Trends Plant Sci* 23(5):451–466
18. Fahlgren N, Gehan MA, Baxter I (2015) Lights, camera, action: high-throughput plant phenotyping is ready for a close-up. *Curr Opin Plant Biol* 24:93–99
19. Liebisch F, Kirchgessner N, Schneider D, Walter A, Hund A (2015) Remote, aerial phenotyping of maize traits with a mobile multi-sensor approach. *Plant Methods* 11(1):9
20. White JW, Conley MM (2013) A flexible, low-cost cart for proximal sensing. *Crop Sci* 53(4):1646–1649
21. Andrade-Sanchez P, Gore MA, Heun JT, Thorp KR, Carmo-Silva AE, French AN, et al (2014) Development and evaluation of a field-based high-throughput phenotyping platform. *Funct Plant Biol* 41(1):68–79
22. Sankaran S, Khot LR, Espinoza CZ, Jarolmasjed S, Sathuvalli VR, Vandemark GJ, et al (2015) Low-altitude, high-resolution aerial imaging systems for row and field crop phenotyping: a review. *Eur J Agron* 70:112–123
23. Yang G, Liu J, Zhao C, Li Z, Huang Y, Yu H, et al (2017) Unmanned aerial vehicle remote sensing for field-based crop phenotyping: current status and perspectives. *Front Plant Sci* 8: 1111
24. Tattaris M, Reynolds MP, Chapman SC (2016) A direct comparison of remote sensing approaches for high-throughput phenotyping in plant breeding. *Front Plant Sci* 7:1131
25. Kefauver SC, Vicente R, Vergara-Díaz O, Fernandez-Gallego JA, Kerfal S, Lopez A, et al (2017) Comparative UAV and field phenotyping to assess yield and nitrogen use efficiency in hybrid and conventional barley. *Front Plant Sci* 8:1733
26. Madec S, Baret F, De Solan B, Thomas S, Dutartre D, Jezequel S, et al (2017) High-throughput phenotyping of plant height: comparing unmanned aerial vehicles and ground LiDAR estimates. *Front Plant Sci* 8: 2002
27. Yu K, Anderegg J, Mikaberidze A, Karisto P, Mascher F, McDonald BA, et al (2018) Hyperspectral canopy sensing of wheat *Sep-*toria tritici** blotch disease. *Front Plant Sci* 9: 1195
28. Yuan W, Li J, Bhatta M, Shi Y, Baenziger P, Ge Y (2018) Wheat height estimation using LiDAR in comparison to ultrasonic sensor and UAS. *Sensors* 18(11):3731
29. Sun S, Li C, Paterson AH, Jiang Y, Xu R, Robertson JS, et al (2018) In-field high throughput phenotyping and cotton plant growth analysis using LiDAR. *Front Plant Sci* 9:16
30. Wang X, Singh D, Marla S, Morris G, Poland J (2018) Field-based high-throughput phenotyping of plant height in sorghum using different sensing technologies. *Plant Methods* 14(1):53
31. Shi Y, Thomasson JA, Murray SC, Pugh NA, Rooney WL, Shafian S, et al (2016) Unmanned aerial vehicles for high-throughput phenotyping and agronomic research. *PloS One* 11(7):e0159781
32. Watanabe K, Guo W, Arai K, Takanashi H, Kajiya-Kanegae H, Kobayashi M, et al (2017) High-throughput phenotyping of sorghum plant height using an unmanned aerial vehicle and its application to genomic prediction modeling. *Front Plant Sci* 8:421
33. Han L, Yang G, Yang H, Xu B, Li Z, Yang X (2018) Clustering field-based maize phenotyping of plant-height growth and canopy spectral dynamics using a UAV remote-sensing approach. *Front Plant Sci* 9:1638
34. Li J, Shi Y, Veeranampalayam-Sivakumar AN, Schachtman DP (2018) Elucidating sorghum biomass, nitrogen and chlorophyll contents with spectral and morphological traits derived from unmanned aircraft system. *Front Plant Sci* 9:1406.
35. Pugh N, Horne DW, Murray SC, Carvalho G, Malambo L, Jung J, et al (2018) Temporal estimates of crop growth in sorghum and maize breeding enabled by unmanned aerial systems. *Plant Phenome J* 1(1):1–10
36. Spindel JE, Dahlberg J, Colgan M, Hollingsworth J, Sievert J, Staggenborg SH, et al (2018) Association mapping by aerial drone reveals 213 genetic associations for Sorghum bicolor biomass traits under drought. *BMC Genomics* 19(1):679

37. Chapman SC, Merz T, Chan A, Jackway P, Hrabar S, Dreccer MF, et al (2014) Pheno-Copter: a low-altitude, autonomous remote-sensing robotic helicopter for high-throughput field-based phenotyping. *Agronomy* 4(2):279–301
38. Guo W, Zheng B, Potgieter AB, Diot J, Watanabe K, Noshita K, et al (2018) Aerial imagery analysis—quantifying appearance and number of sorghum heads for applications in breeding and agronomy. *Front Plant Sci* 9: 1544
39. Xu R, Li C, Paterson AH, Jiang Y, Sun S, Robertson JS (2018) Aerial images and convolutional neural network for cotton bloom detection. *Front Plant Sci* 8:2235
40. Haghighattalab A, Pérez LG, Mondal S, Singh D, Schinstock D, Rutkoski J, et al (2016) Application of unmanned aerial systems for high throughput phenotyping of large wheat breeding nurseries. *Plant Methods* 12(1):35
41. Zaman-Allah M, Vergara O, Araus J, Tarekegne A, Magorokosho C, Zarco-Tejada P, et al (2015) Unmanned aerial platform-based multi-spectral imaging for field phenotyping of maize. *Plant Methods* 11(1):35
42. Condorelli GE, Maccaferri M, Newcomb M, Andrade-Sanchez P, White JW, French AN, et al (2018) Comparative aerial and ground based high throughput phenotyping for the genetic dissection of NDVI as a proxy for drought adaptive traits in durum wheat. *Front Plant Sci* 9:893
43. Potgieter AB, George-Jaeggli B, Chapman SC, Laws K, Suárez Cadavid LA, Wixted J, et al (2017) Multi-spectral imaging from an unmanned aerial vehicle enables the assessment of seasonal leaf area dynamics of sorghum breeding lines. *Front Plant Sci* 8:1532
44. Kyratzis AC, Skarlatos DP, Menexes GC, Vamvakousis VF, Katsiotis A (2017) Assessment of vegetation indices derived by UAV imagery for durum wheat phenotyping under a water limited and heat stressed Mediterranean environment. *Front Plant Sci* 8: 1114
45. Hassan MA, Yang M, Rasheed A, Jin X, Xia X, Xiao Y, et al (2018) Time-series multispectral indices from unmanned aerial vehicle imagery reveal senescence rate in bread wheat. *Remote Sensing* 10(6):809
46. Vergara-Díaz O, Zaman-Allah MA, Masuka B, Hornero A, Zarco-Tejada P, Prasanna BM, et al (2016) A novel remote sensing approach for prediction of maize yield under different conditions of nitrogen fertilization. *Front Plant Sci* 7:666
47. Jay S, Baret F, Dutartre D, Malatesta G, Héno S, Comar A, et al (2019) Exploiting the centimeter resolution of UAV multispectral imagery to improve remote-sensing estimates of canopy structure and biochemistry in sugar beet crops. *Remote Sens Environ* 231: 110898
48. Verger A, Vigneau N, Chéron C, Gilliot JM, Comar A, Baret F (2014) Green area index from an unmanned aerial system over wheat and rapeseed crops. *Remote Sens Environ* 152:654–664
49. Khan Z, Rahimi-Eichi V, Haefele S, Garnett T, Miklavcic SJ (2018) Estimation of vegetation indices for high-throughput phenotyping of wheat using aerial imaging. *Plant Methods* 14(1):20
50. Ludovisi R, Tauro F, Salvati R, Khoury S, Mugnoz Scarascia G, Harfouche A (2017) UAV-based thermal imaging for high-throughput field phenotyping of black poplar response to drought. *Front Plant Sci* 8:1681
51. Acquaah G (2009) Principles of plant genetics and breeding. Wiley, New York
52. Thorp K, Thompson A, Harders S, French A, Ward R (2018) High-throughput phenotyping of crop water use efficiency via multispectral drone imagery and a daily soil water balance model. *Remote Sens* 10(11):1682
53. Allen RG, Pereira LS, Raes D, Smith M, et al (1998) Crop evapotranspiration: guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. Fao, Rome. 300(9):D05109
54. Falconer DS, Mackay T (1996) Introduction to quantitative genetics. Oliver And Boyd, Edinburgh, London
55. Sankaran S, Khot LR, Carter AH (2015) Field-based crop phenotyping: Multispectral aerial imaging for evaluation of winter wheat emergence and spring stand. *Comput Electron Agricult* 118:372–379
56. Rutkoski J, Poland J, Mondal S, Autrique E, Pérez LG, Crossa J, et al (2016) Canopy temperature and vegetation indices from high-throughput phenotyping improve accuracy of pedigree and genomic selection for grain yield in wheat. *G3 Genes Genomes Genetics* 6:g3–116.
57. Elias AA, Rabbi I, Kulakow P, Jannink JL (2018) Improving genomic prediction in cassava field experiments using spatial analysis. *G3 Genes Genomes Genetics* 8(1):53–62
58. Xavier A, Hall B, Hearst AA, Cherkauer KA, Rainey KM (2017) Genetic architecture of phenomic-enabled canopy coverage in Glycine max. *Genetics* 206:p. genetics–116

59. Sun J, Rutkoski JE, Poland JA, Crossa J, Janink JL, Sorrells ME (2017) Multitrait, random regression, or simple repeatability model in high-throughput phenotyping data improve genomic prediction for wheat grain yield. *Plant Genome* 10(2): plantgenome2016–11
60. Cabrera-Bosquet L, Crossa J, von Zitzewitz J, Serret MD, Luis Araus J (2012) High-throughput phenotyping and genomic selection: the frontiers of crop breeding converge. *J Integr Plant Biol* 54(5):312–320
61. Krause MR, González-Pérez L, Crossa J, Pérez-Rodríguez P, Montesinos-López O, Singh RP, et al (2019) Hyperspectral reflectance-derived relationship matrices for genomic prediction of grain yield in wheat. *G3 Genes Genomes Genetics* 9(4): 1231–1247
62. Juliana P, Montesinos-López OA, Crossa J, Mondal S, Pérez LG, Poland J, et al (2019) Integrating genomic-enabled prediction and high-throughput phenotyping in breeding for climate-resilient bread wheat. *Theoret Appl Genetics* 132(1):177–194
63. Jarquín D, Howard R, Xavier A, Das Choudhury S (2018) Increasing predictive ability by modeling interactions between environments, genotype and canopy coverage image data for soybeans. *Agronomy* 8(4):51
64. Amani I, Fischer R, Reynolds M (1996) Canopy temperature depression association with yield of irrigated spring wheat cultivars in a hot climate. *J Agronomy Crop Sci* 176(2): 119–129
65. Aguete FM, Trachsel S, Pérez LG, Burgueño J, Crossa J, Balzarini M, et al (2017) Use of hyperspectral image data outperforms vegetation indices in prediction of maize yield. *Crop Sci* 57(5):2517–2524
66. Montesinos-López OA, Montesinos-López A, Crossa J, los Campos G, Alvarado G, Suchismita M, et al (2017) Predicting grain yield using canopy hyperspectral reflectance in wheat breeding data. *Plant Methods* 13(1):4
67. Roorkiwal M, Jarquín D, Singh MK, Gaur PM, Bharadwaj C, Rathore A, et al (2018) Genomic-enabled prediction models using multi-environment trials to estimate the effect of genotype $\times$ environment interaction on prediction accuracy in chickpea. *Sci Rep* 8(1):11701
68. Sukumaran S, Crossa J, Jarquín D, Reynolds M (2017) Pedigree-based prediction models with genotype $\times$ environment interaction in multi-environment trials of CIMMYT wheat. *Crop Sci* 57(4):1865–1880
69. Jarquín D, Lemes da Silva C, Gaynor RC, Poland J, Fritz A, Howard R, et al (2017) Increasing genomic-enabled prediction accuracy by modeling genotype $\times$ environment interactions in Kansas Wheat. *Plant Genome* 10(2):plantgenome2016–12
70. Montesinos-López A, Montesinos-López OA, Cuevas J, Mata-López WA, Burgueño J, Mondal S, et al (2017) Genomic Bayesian functional regression models with interactions for predicting wheat grain yield using hyperspectral image data. *Plant Methods* 13(1):62
71. Davis C (1962) The norm of the Schur product operation. *Numer Math* 4(1):343–344
72. Montesinos-López A, Montesinos-López OA, Campos G, Crossa J, Burgueño J, Luna-Vazquez FJ (2018) Bayesian functional regression as an alternative statistical analysis of high-throughput phenotyping data of modern agriculture. *Plant Methods* 14(1):46
73. Shavrukov Y, Kurishbayev A, Jatayev S, Shvidchenko V, Zotova L, Koekemoer F, et al (2017) Early flowering as a drought escape mechanism in plants: How can it aid wheat production? *Front Plant Sci* 8:1950
74. Zhang J, Yang C, Song H, Hoffmann WC, Zhang D, Zhang G (2016) Evaluation of an airborne remote sensing platform consisting of two consumer-grade cameras for crop identification. *Remote Sens* 8(3):257
75. Berger B, Parent B, Tester M (2010) High-throughput shoot imaging to study drought responses. *J Exper Botany* 61(13): 3519–3528
76. Campbell MT, Knecht AC, Berger B, Brien CJ, Wang D, Walia H (2015) Integrating image-based phenomics and association analysis to dissect the genetic architecture of temporal salinity responses in rice. *Plant Physiol* 168(4):1476–1489.
77. Apiolaza LA, Gilmour AR, Garrick DJ (2000) Variance modelling of longitudinal height data from a *Pinus radiata* progeny test. *Can J For Res* 30(4):645–654
78. Apiolaza LA, Garrick DJ (2001) Analysis of longitudinal data from progeny tests: some multivariate approaches. *Forest Sci* 47(2): 129–140
79. de Souza Marçal T, Salvador FV, da Silva AC, Machado JC, Carneiro PCS, et al (2018) Genetic insights into elephantgrass persistence for bioenergy purpose. *PloS One* 13(9):e0203818
80. Wu WR, Li WM, Tang DZ, Lu HR, Worland A (1999) Time-related mapping of quantitative trait loci underlying tiller number in rice. *Genetics* 151(1):297–303

81. Yan J, Zhu J, He C, Benmoussa M, Wu P (1998) Molecular dissection of developmental behavior of plant height in rice (*Oryza sativa* L.). *Genetics* 150(3):1257–1265
82. Würschum T, Liu W, Busemeyer L, Tucker MR, Reif JC, Weissmann EA, et al (2014) Mapping dynamic QTL for plant height in triticale. *BMC Genetics* 15(1):59
83. Moore CR, Johnson LS, Kwak IY, Livny M, Broman KW, Spalding EP (2013) High-throughput computer vision introduces the time axis to a quantitative trait map of a plant growth response. *Genetics* 195:p. genetics–113
84. Paine CT, Marthews TR, Vogt DR, Purves D, Rees M, Hector A, et al (2012) How to fit nonlinear plant growth models and calculate growth rates: an update for ecologists. *Methods Ecol Evol* 3(2):245–256
85. Ma CX, Casella G, Wu R (2002) Functional mapping of quantitative trait loci underlying the character process: a theoretical framework. *Genetics* 161(4):1751–1762
86. Wu R, Ma CX, Zhao W, Casella G (2003) Functional mapping for quantitative trait loci governing growth rates: a parametric model. *Physiol Genomics* 14(3):241–249
87. Wu R, Lin M (2006) Functional mapping – how to map and study the genetic architecture of dynamic complex traits. *Nat Rev Genetics* 7(3):229
88. Das K, Li J, Wang Z, Tong C, Fu G, Li Y, et al (2011) A dynamic model for genome-wide association studies. *Human Genetics* 129(6):629–639
89. Cui Y, Zhu J, Wu R (2006) Functional mapping for genetic control of programmed cell death. *Physiol Genomics* 25(3):458–469
90. He Q, Berg A, Li Y, Vallejos CE, Wu R (2010) Mapping genes for plant structure, development and evolution: functional mapping meets ontology. *Trends Genetics* 26(1):39–46
91. Bac-Molenaar JA, Vreugdenhil D, Granier C, Keurentjes JJ (2015) Genome-wide association mapping of growth dynamics detects time-specific and general quantitative trait loci. *J Exper Botany* 66(18):5567–5580
92. Campbell MT, Du Q, Liu K, Brien CJ, Berger B, Zhang C, et al (2017) A comprehensive image-based phenomic analysis reveals the complex genetic architecture of shoot growth dynamics in rice (*Oryza sativa*). *Plant Genome* 10(2):plantgenome2016–07
93. Schaeffer L (1994) Random regressions in animal models for test-day production in dairy cattle. In: *World Congress of Genetics Applied Livestock Production*, vol 18. pp 443–446
94. Schnyder U, Hofer A, Labroue F, Künzi N (2001) Genetic parameters of a random regression model for daily feed intake of performance tested French Landrace and Large White growing pigs. *Genet Sel Evol* 33(6):635
95. Luo P, Yang R, Yang N (2007) Estimation of genetic parameters for cumulative egg numbers in a broiler dam line by using a random regression model. *Poultry Sci* 86(1):30
96. Baldi F, Albuquerque L, Alencar M (2010) Random regression models on Legendre polynomials to estimate genetic parameters for weights from birth to adult age in Canchim cattle. *J Animal Breeding Genetics* 127(4):289–299
97. Howard JT, Jiao S, Tiezzi F, Huang Y, Gray KA, Maltecca C (2015) Genome-wide association study on legendre random regression coefficients for the growth and feed intake trajectory on Duroc Boars. *BMC Genetics* 16(1):59
98. Kirkpatrick M, Lofsvold D, Bulmer M (1990) Analysis of the inheritance, selection and evolution of growth trajectories. *Genetics* 24(4):979–993
99. Schaeffer L (2004) Application of random regression models in animal breeding. *Livest Prod Sci* 86(1–3):35–45
100. Meyer K, Hill WG (1997) Estimation of genetic and phenotypic covariance functions for longitudinal or ‘repeated’ records by restricted maximum likelihood. *Livest Prod Sci* 47(3):185–200
101. Pool M, Meuwissen T (2000) Reduction of the number of parameters needed for a polynomial random regression test day model. *Livest Prod Sci* 64(2–3):133–145
102. Campbell M, Walia H, Morota G (2018) Utilizing random regression models for genomic prediction of a longitudinal trait derived from high-throughput phenotyping. *Plant Direct* 2(9):e00080
103. Baba T, Momen M, Campbell MT, Walia H, Morota G (2020) Multi-trait random regression models increase genomic prediction accuracy for a temporal physiological trait derived from high-throughput phenotyping. *PloS One* 15(2):e0228118

104. Habier D, Fernando R, Dekkers JC (2007) The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177(4):2389–2397
105. Jannink JL (2010) Dynamics of long-term genomic selection. *Genet Sel Evol* 42(1):35
106. Iwata H, Hayashi T, Tsumura Y (2011) Prospects for genomic selection in conifer breeding: a simulation study of *Cryptomeria japonica*. *Tree Genet Genomes* 7(4):747–758
107. Yabe S, Ohsawa R, Iwata H (2013) Potential of genomic selection for mass selection breeding in annual allogamous crops. *Crop Sci* 53(1):95–105

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



INDEX

A

- Abiotic stress 4, 20, 138,
165, 169, 229, 270
- Aboveground 37–47, 49, 128, 175, 179
- Agricolae 58, 59, 67
- Arabidopsis 19–21, 46,
78, 224, 283, 285
- Association between co-expression network and
phenotyping data 261–266
- Automated data collection 162

B

- Breeding 67, 71, 77, 119, 136,
161, 171, 213, 225, 236, 269

C

- Camelina sativa*, 25–36
- Camera calibration 88, 177
- Chlorophyll fluorescence 4, 21, 213–219
- Color patterning 77, 78, 81, 82, 84
- Color phenotyping 77–84
- Common bean 7, 101, 102, 113, 114
- Construction 148, 191–210,
262, 263, 283
- Continuous color distribution 80–82
- Correlated measures 78, 160, 274, 277, 278

D

- Data integration 278, 290
- Design 26, 35, 58, 100,
121, 150, 165, 172, 191, 215, 225, 237
- Digital camera 38, 80, 271
- Drought 7, 137, 138,
165, 223–232, 283
- Drought marker genes 229–231

E

- Excavated root system 123, 124, 126, 128, 130
- Experimental design v, 26, 28, 35,
57–67, 100, 166, 225, 227, 237, 253
- Experimental workflow 160, 179, 180,
185, 237, 239, 240, 250, 254, 255, 257

F

- Feature extraction 71–74,
120–122, 125, 129
- Field-based phenotyping 171
- Field phenotyping 213–219
- Field scanner 214–217

G

- GC-high-resolution (HR)-time-of-flight (TOF)-MS
(GC-HR-TOF-MS) 237, 239, 240, 242
- Gene co-expression network 261–266
- Genetic gain 270, 275–278, 280, 290
- Genomes2Fields (G2F) 58, 64–67
- Greenhouse 26–29, 36–38,
49–55, 64–67, 123, 162, 165, 171, 277

H

- High-throughput image acquisition 13, 21, 79
- High-throughput phenotyping (HTP) 11–17,
19–23, 25–36, 77–84, 87, 137, 161, 167, 171,
214, 269–291
- High-throughput plant phenotyping (HTPP) v,
57–67, 136
- High-throughput screening (HTS) 3–8,
13, 20, 21
- Hydroponics 19–23,
101, 105, 106, 108

I

- Image analysis v, 4, 6–8, 15–17,
20, 26, 28, 30–34, 44, 45, 47, 79, 120, 160, 161,
172, 186
- Image data 26, 172,
180, 185, 187, 214, 281
- Image measurement 87–93
- Image processing 6, 16, 30,
31, 38, 43, 47, 55, 72–75, 84, 87, 91, 120
- Imaging 4, 12, 29, 38, 64,
71, 78, 97, 119, 139, 162, 176, 197, 213, 273
- Inflorescence 121, 129

L

- Longitudinal trait 282–286

M

Marchantia polymorpha 11–17
 Metabolomics v, 235, 236, 242, 254, 255
 Module discovery 263, 266
 Molecular imaging v, 99
 Morphological parameters 13, 32

N

Nitrogen 19–23, 107, 108, 139, 226–228, 231, 238, 241, 242, 256, 273

P

Panicle 123, 124, 126–129
 Phenomics v, 25–36, 119, 135–152, 159–170, 172, 290
 Phenotypes v, 13, 20, 37, 38, 45, 49, 57, 58, 64, 75, 77–84, 87, 137, 140, 159–161, 163–166, 168, 170, 175, 213, 227, 235–259, 261, 262, 264, 270, 274, 275, 282, 283, 285, 287, 290, 291
 Phenotyping 4, 12, 20, 25, 37, 55, 57, 75, 77, 87, 99, 119, 136, 161, 171, 191, 213, 227, 235, 262, 270
 Photosystem II (PSII) 4, 138, 214, 215, 219
 Plant architecture 38, 171–188, 218
 Plant phenotyping 12, 17, 37, 49–55, 87, 99, 119, 188, 191–210, 213, 235
 Positron emission tomography (PET) v, 97–116

Q

Quantitative genetics v, 168, 169, 269–291

Quantitative reverse transcription-polymerase chain reaction (qRT-PCR) 225, 226, 228–230
 Quantum efficiency 214, 219

R

Raspberry Pis (Pis) 49–55, 180, 196, 200, 204–207, 209, 210
 Remote phenotyping 139, 168, 214, 271
 RGB images 273
 RGB (red, green, blue) 4–6, 21, 22, 30, 34, 75, 79–81, 163, 175, 186, 192, 198–200, 215, 272, 273
 Robotics 191, 199, 200
 Root 20, 82, 87, 101, 120, 149, 172, 207, 224, 242, 275
 Row crops 172, 173

S

Seed phenotyping 75
 Senescence 3, 7, 139, 224
 Shape-independent color quantification 78
 Sorghum 120, 121, 123, 124, 126–129, 172, 176, 177, 187, 193, 214, 223–232, 272, 277
 Stay-green 3–8, 139, 224
 Sub-canopy 191–210

T

3D reconstruction 49, 51, 53–55, 125, 128, 130, 178, 185

X

X-ray tomography 119–130