Uncovering the genetic basis of seed size and root gravitropism using high-throughput phenotyping of *Arabidopsis thaliana*

By

Candace R. Moore

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Botany)

at the

UNIVERSITY OF WISCONSIN-MADISON

2013

Date of final oral examination: 12/02/13

The dissertation is approved by the following members of the Final Oral Committee:

David A. Baum, Professor, Botany
Simon Gilroy, Professor, Botany
Patrick Masson, Professor, Genetics
Edgar P. Spalding, Professor, Botany
Brian S. Yandell, Professor, Horticulture and Statistics

Acknowledgments

I would like to thank Kelsey Rudd and David Gronwall for assistance in planting, maintaining, and harvesting seed for material used in RIL1 and NIL datasets. David Gronwall and Laura Vaughn were especially helpful with discussions that aided my early understanding of the principles of QTL analysis. Thanks to Nathan Miller for writing the image analysis software to measure seed size traits analyzed in Chapter 2.

A special thanks is due to Logan Johnson, who in addition to customizing software for root tip angle measurements on stacks of camera images, worked diligently with HTCondor to enable QTL mapping of many phenotypes at once over distributing computing networks, crucial to the analyses presented in Chapter 3. Thanks also to Miron Livny, Bill Taylor, and other members of the HTCondor team for the use of their software as well as their desire to improve HTCondor's ability to manage jobs requiring the R software suite as well as memory-intensive tasks. In addition, our collaboration with Karl Broman and Il-Youp Kwak both aided my understanding of QTL analysis as well as significantly contributed to the publication of Chapter 3.

Thank you to my committee members, Jean Michel Ané, David A. Baum, Patrick Masson, and Brian S. Yandell for their ideas and constructive comments on the work presented here.

I would especially like to thank Edgar P. Spalding for his unwavering support, without which I could not be here today. His contagious love of science, tremendous ability to provide financial support for the lab, and desire to provide an environment where exchange of ideas

and continuous learning occurs makes him both a wonderful P.I. and friend. Thanks also to all the members of the Spalding lab for sharing their ideas, expertise, and friendship.

Last, I thank my family for their continued encouragement and support. I thank my husband, Matthew Moore, for his constant support, love, and willingness to move far away from North Carolina while I pursued my degree, and my dear cat Bob whose sweetness and antics kept me both happy and amused during these years. I thank my parents who always encouraged me to follow my dreams and my grandmother, Betty Horak, who always believed in me.

Publications Arising from This Thesis

MOORE, C. R., D. S. GRONWALL, N. D. MILLER and E. P. SPALDING, 2013a Mapping quantitative trait loci affecting *Arabidopsis thaliana* seed morphology features extracted computationally from images. G3: Genes|Genomes|Genetics **3:** 109-118 (Chapter 2).

MOORE, C. R., L. S. JOHNSON, I.-Y. KWAK, M. LIVNY, K. W. BROMAN, and E. P. SPALDING, 2013b High-throughput computer vision introduces the time axis to a quantitative trait map of a plant growth response. Genetics **195:** 1077-1086 (Chapter 3).

KWAK, I.-Y., C. R. MOORE, E. P. SPALDING, and K. W. BROMAN, Mapping quantitative trait loci underlying function-valued phenotypes. Manuscript in process (Chapter 3).

Preface

The opportunity to improve plant species important for global food security by understanding their genetics and physiology was a major factor in my desire to study plant biology. Through my undergraduate work on inositol-1,4,5-trisphosphate, a signaling molecule involved in plant responses to stimuli such as drought and changes in gravity, I was introduced to Edgar Spalding, whose laboratory studied the characteristics of proteins involved in the movement of auxin, a major plant hormone that affects cellular elongation, and thus organ development. When I began my graduate work, members of the laboratory including Nathan Miller, Brian Parks, and Daniel Lewis had begun using image analysis techniques that quantified processes such as hypocotyl elongation and root gravitropism. These early algorithms proved extremely useful, but were limited in terms of the amount of user manipulation required as well as by the computational resources available. However, these algorithms did provide detailed analyses of seedlings' light responses and root growth (MILLER *et al.* 2007), such as how different cellular pools of cryptochrome affect seedling photomorphogenesis (WU and SPALDING 2007) and the role of the ABCB family of proteins in auxin transport (LEWIS *et al.* 2007).

Other work in the lab by Tessa Durham Brooks and Angela Elwell focused on the effects of maternal growth environment on seed size distribution (ELWELL *et al.* 2011) and the impact of seed size on the root gravitropic response (DURHAM BROOKS *et al.* 2010). The fact that we as scientists needed to pay close attention to the growing conditions under which our biological material was produced was quickly becoming apparent. Also around this time, I was introduced to Laura Vaughn, then a graduate student in Patrick Masson's laboratory, who was using a

recombinant inbred line population to measure the genetic basis of root skewing in Arabidopsis and generously shared seed stock with me. This led to my collaborating with Nathan Miller to produce custom software that precisely and accurately measures seed parameters including area, length, and width. Our aim was to use the recombinant inbred line population to perform a high-throughput quantitative trait loci analysis of seed size in Arabidopsis with virtually no user manipulation required.

Subsequently, the seeds given to us from Laura Vaughn were used as progenitors of a second replicate of the recombinant line population, reared in the UW-Madison Biotron. A set of near isogenic lines derived from the same parental accessions was also grown in the same manner, and all seed progeny was imaged in high resolution. The phenotype values derived from these images were used to perform quantitative trait loci (QTL) analyses, the results of which are presented in Chapter 2.

As the large and numerous image files of seeds were being gathered, the need for a more advanced computing infrastructure became obvious. Under the guidance of Botany Information Technology Specialist Thomas Maher, I created and maintained Linux-based servers for use of the Spalding lab. For the first time, data could be viewed and shared from any computer in the laboratory, and regularly scheduled backups of all members' data occurred. Over time, Logan Johnson built and expanded on this model, building servers with increased storage capacity and faster data transfer speeds. The advances made to our computing infrastructure were crucial to our ability to automatically and rapidly transfer files to

the Center for High-Throughput Computing, where our feature extraction and QTL analyses were performed.

Now that we had high quality seed stocks of three mapping populations from which genetic factors underlying the natural variation in seed size could be identified, we wondered if the same gene or genes involved in these traits could also affect the natural variation of the root gravitropic response. Tessa Durham Brooks had shown that for seeds from the wild type accession Columbia, the size of the seed determined the time course of the seedling root's gravitropic response (DURHAM BROOKS *et al.* 2010). Perhaps we could find genetic factors that would allow us to predict how a root would respond to a reorientation based partly on the size of the seed from which it grew.

From there I began collaboration with Logan Johnson, who modified an algorithm written by Nathan Miller so that root tip angles could be measured from camera image stacks of gravitroping roots using principal components analysis. We imaged thousands of roots from our three inbred populations, partnering with the HTCondor team to perform feature extraction over a distributed computing network. Although we did not find common loci that control both root gravitropism and seed size, we were able to identify and characterize several loci that contribute to the variation in the root gravitropic response, using tip angle values at each time point as the phenotype. We shared our phenotype data with Karl Broman and Il-Youp Kwak, who devised a method of trait mapping that controls for the search across time. The results of both methods of QTL analysis of root gravitropism are discussed in Chapter 3.

The remarkable outcome from this project was our ability to visualize the timing of action of each significant locus, a phenomenon which, to my knowledge, has never been seen.

Throughout my thesis work, I have had the opportunity to learn much more than I could have imagined. Much of the credit goes to Edgar Spalding, who brought together the incredible team of researchers necessary to make this collaborative work a success. I am extremely grateful for the interactions I have had with statisticians, computer scientists, and plant biologists, and hopeful that together, we will be able to address the phenotyping challenges of the future.

Table of Contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

Appendix

List of Tables

Chapter 1

Chapter 2

Abstract

One of the major goals in plant biology is to resolve the function of every gene within a plant's genome and to measure the natural variation present in its alleles. Phenotyping, the process of quantifying observable traits, is one method to assess the underlying nature of the genetic factors that control the traits. Fully automated, high-throughput phenotyping platforms that provide fast, accurate measurements of mapping populations are enabling new insights into basic physiological functions important for continued global food security. Key to these endeavors are sophisticated robotic imaging devices, image analysis algorithms, smart database design, and knowledge-sharing with the research community. This thesis describes the creation of high-throughput systems for measuring seed size and root gravitropism, traits fundamental to agriculture, and identifies genomic regions that contribute to the variation in these traits. Flatbed scanners were used to create digital images of thousands of seeds at once, which were then analyzed with computer vision to quantify area, length, and width. Similarly, infrared-sensitive cameras captured the growth of roots in response to a reorientation of the gravity vector at high temporal resolution, generating a stack of images from which the development of tip angle can be measured using principal components analysis. Determining these feature values in members of two genetic replicates of the L*er* x Cvi recombinant inbred line population and a near-isogenic population derived from the same accessions permitted quantitative trait loci analyses that identified numerous genetic regions involved in these traits. Genetic loci that controlled variation across only one axis of the seed were located by calculation of width and length, as opposed to solely seed area. For the root gravitropism

phenotype, the high spatial and temporal resolution combined with two methods of genetic mapping revealed the timing of action of the loci involved in this response. The images, algorithms, measurements, and trait mapping results are provided for download and further study to the plant biology community. Utilization of these phenotyping platforms to measure natural variation present in other mapping populations grown in a variety of environmental conditions will allow a fuller understanding of these important physiological processes.

Chapter 1:

The root of the matter: Uncovering the secrets of natural variation by high-throughput phenotyping

Abstract

The future of crop improvement relies on scientists' ability to accurately measure biologically meaningful phenotypes and relate these traits to the particular genetic alleles that control them. Numerous mapping populations and natural accessions exist that remain to be phenotyped in a variety of conditions to explore the complex relationships between genetics and the environment. Much attention is now being paid to the growth and development of roots due to their important roles in crop establishment, sugar storage, and water and nutrient uptake. Slow, manual, and destructive methods of imaging root system architecture are paving way to automated, high-throughput systems necessary to catalog the wide diversity and plasticity of plant responses. These platforms help alleviate the phenotyping bottleneck, but must be combined with reliable recording of environmental parameters, appropriate algorithms to precisely quantify trait data, and ontology-centered database structures that allow for organization and searching of the results. Also crucial is the need for public repositories of raw images from QTL studies so that the data can be both independently verified and explored for novel genotype x phenotype interactions. These advancements will promote new understandings of gene function and hopefully improve food security across the globe.

Introduction

"Accurate and minute measurement seems to the non-scientific imagination a less lofty and dignified work than looking for something new. But nearly all the grandest discoveries of science have been but the rewards of accurate measurement and patient long-continued labour in the minute sifting of numerical results."

- Baron William Thomson Kelvin

Presidential inaugural address to the General Meeting of the British Association, Edinburgh (August 2, 1871)

As the population of the earth continues to increase, the need for global food security becomes increasingly apparent. Increases in yield driven by the Green Revolution have started to stabilize (RAY *et al.* 2012), and the next revolution in agriculture will be dependent on a deeper understanding of the genetic makeup of crop species and the alleles which can confer enhanced abilities to mine nutrients from the soil, resist drought and other abiotic stresses, and defend against attacks from pathogens. Understanding the role that particular alleles play in determining the phenotype of a plant has become a major challenge for plant biology. Agricultural phenotypes important for crop survival and yield include photosynthetic capacity, leaf area, seed size, and root system architecture (RSA). Root phenotypes are perhaps the most difficult to study simply because their development progresses within a hidden realm of soil,

and they exhibit a great level of plasticity in their responses to it (FITZ GERALD *et al.* 2006; GRUBER *et al.* 2013).

Too often, the rate limiting step in identifying the genes underlying these phenotypes is the ability to quickly and accurately measure the traits of interest. Early studies of natural variation of root traits were often performed in field settings or in soil or sand-filled containers, requiring the destruction of the sample in order to take manual measurements of the roots (KAMOSHITA *et al.* 2002; LAFITTE *et al.* 2001; LI *et al.* 2005; ZHENG *et al.* 2003). Image analysis programs using computer vision to quantify features from digital images, usually from scanned pictures of washed root systems, are somewhat of an improvement, but due to the destructive procedure required for root imaging, are not conducive to temporal analysis of root behavior, as well as still being time-intensive for the researcher (DOUGLAS 1991; FITA *et al.* 2007; ZHU *et al.* 2006; ZHU *et al.* 2005). Non-destructive, automated phenotyping platforms combined with computer vision to measure the traits of interest help to overcome this bottleneck and facilitate temporal studies of growth. When performed on appropriate mapping populations, automated phenotyping can aid in assigning function to previously uncharacterized genes. Growing plants in liquid culture (COTSAFTIS *et al.* 2011; MIYAMOTO *et al.* 2001) or on a two-dimensional surface such as agar or paper (BENGOUGH *et al.* 2004; FITA *et al.* 2008; HUND *et al.* 2009) provides a non-destructive method for phenotyping plants for QTL studies at discrete time intervals. For 3D reconstructions of root architecture, samples can be grown in clear gel-filled tubes (FANG *et al.* 2009; FANG *et al.* 2013; FITA *et al.* 2010; IYER-PASCUZZI *et al.* 2010) or a novel type of clear soil

(DOWNIE *et al.* 2012; YANG *et al.* 2013). These growth systems further our understanding of root behavior by enabling digital imaging without spatially constraining RSA.

This review will especially consider the role automated phenotyping can play in understanding the genetic basis of important traits, focusing on the diversity of mapping populations, phenotyping systems, software for RSA analysis, and data management structures necessary for robust phenotyping workflows (Figure 1). For more general reviews, see de Souza (2010), Normanly (2012), and Sozzani and Benfey (2011). Interdisciplinary collaboration between plant and computer scientists is crucial for full integration of these tools. As these systems improve, perhaps grand discoveries will be made without the "patient long-continued labour" and "minute sifting of numerical results" that Baron William Thomson Kelvin endured.

Mapping Populations

The purpose of the phenotyping platforms and software to be discussed in this review is to advance the ability to acquire high quality phenotype data of root traits in organisms such as Arabidopsis, rice, and maize. In order to identify the genetic loci involved in the variation of these traits, these systems may be used to measure the diversity of mapping populations that are composed of numerous genetically distinct lines in which information has been gathered regarding the differences in DNA sequences. Stock centers that supply these seed stocks, as well as T-DNA insertional mutants, are listed in Table 1. By phenotyping a genetically diverse

Figure 1: Central elements of a successful phenotyping workflow. High-resolution cameras, often robotically controlled, can enable the acquisition of large data sets. Feature extraction can be performed on the resulting images by custom software. Phenotype, genotype, and experimental meta-data must be recorded in relational databases that will allow sharing of data with other members of the research community to inform future experiments.

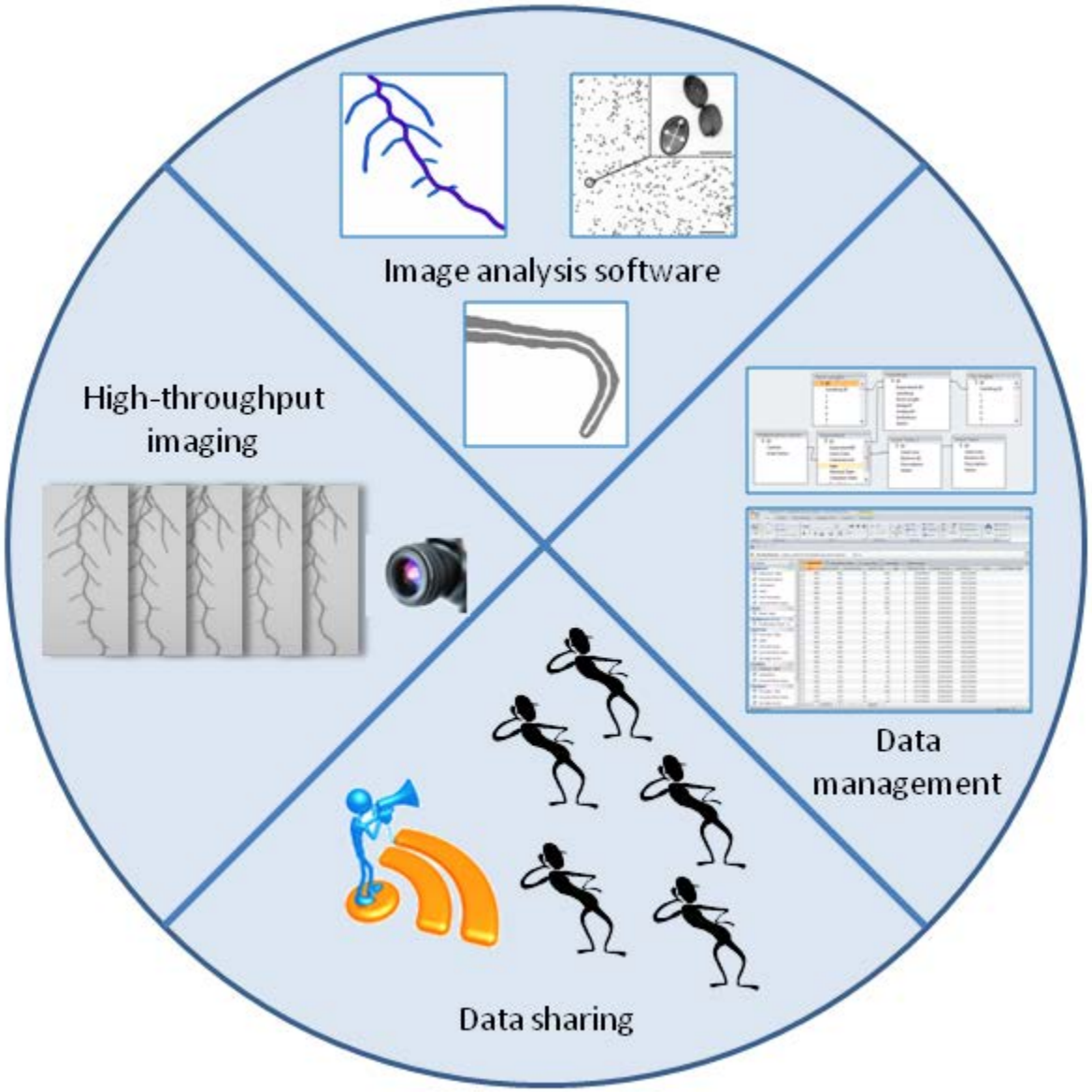


Table 1: Arabidopsis stock centers.

Stock Center	Location	Website
Arabidopsis Biological Resource Center (ABRC)	USA	http://abrc.osu.edu/
Lehle Seeds	USA	http://www.arabidopsis.com/
Nottingham Arabidopsis Stock Centre (NASC)	UK	http://arabidopsis.info/
RIKEN Bioresource Center (BRC)	Japan	http://www.brc.riken.jp
Versailles Arabidopsis Stock Center (VASC)	France	http://publiclines.versailles.inra.fr/

set of lines, one can use statistical mapping approaches in order to find genetic regions that are likely to have positive or negative effects on the trait value. One type of mapping population is a recombinant inbred line (RIL) family, in which two genetically distinct parental accessions have been mated, and then the resulting F2 progeny are self-fertilized and lines maintained through single seed descent until each line is almost completely homozygous (ALONSO-BLANCO *et al.* 1998; LISTER and DEAN 1993; REITER *et al.* 1992). Once these RILs have been genotyped at molecular markers spaced along each chromosome, they can be used for a variety of phenotypic screens in multiple environments to decipher which alleles control the variation in the phenotype in each condition. An increased amount of genetic recombination can be achieved via creation of advanced intercross RILs due to inter-mating of F2 individuals before lines are inbred (BALASUBRAMANIAN *et al.* 2009; DARVASI and SOLLER 1995). Although detection of loci that influence the trait of interest is dependent on allelic differences in the parent lines and the recombination events that occurred during creation of the lines, this type of mapping population also lends itself to studies of epistatic relationships between loci.

Near isogenic lines (NILs) are introgression lines in which a short segment of one parental accession is introgressed into the genome of another parent accession (ESHED and ZAMIR 1995; RHODES *et al.* 1989; SEEVERS *et al.* 1971; TÖRJÉK *et al.* 2008), most commonly through a repeated backcross procedure (PATERSON *et al.* 1990). Although the reduced amount of genetic recombination means a lesser chance of finding large effect QTL, these lines have been shown to be useful in identifying QTL of small effect (KEURENTJES *et al.* 2007). Additionally, NILs can be used as starting material for fine mapping, permitting a narrower region containing a

QTL to be identified (VAUGHN and MASSON 2011). A similar type of mapping population is that of a heterogeneous inbred family (HIF) (TUINSTRA *et al.* 1997). HIFs are derived from NILs that are not completely homozygous, which when self-fertilized, result in progeny that segregate for the non-homozygous loci. These lines can be used to create populations that undergo further recombination events within the non-homozygous interval (HALEY *et al.* 1994). Mapping populations created with these methods enable comparisons of individuals with consistent genetic backgrounds that can be used to find phenotypes controlled by a particular locus.

Although the use of inbred lines derived from two parental accessions have proven useful for QTL mapping studies, these populations are limited in that only alleles which differ between the two parental lines can be identified as controlling a trait of interest. If there is not variation between the two lines at a particular locus, that locus will never be identified by traditional QTL mapping, even if that gene is important to the trait. A new area of research involves sequencing and publishing the full genomes of many Arabidopsis accessions to determine single nucleotide changes between accessions. This work by the 1001 Genomes Project (WEIGEL and MOTT 2009) is enabling genome-wide association studies (GWAS) in which many accessions containing many genetic variants at a particular allele are phenotyped and statistical methods used to identify associations between the trait and the causal variant. At the time of this writing, over 1001 accessions have been sequenced. GWAS has been successfully used in Arabidopsis to identify loci involved in a variety of phenotypes (ATWELL *et al.* 2010), including pathogen resistance (CHAN *et al.* 2011; NEMRI *et al.* 2010), flowering (BRACHI *et*

al. 2010), shade avoidance (FILIAULT and MALOOF 2012), and leaf growth and development (ZHANG *et al.* 2012).

Automated Phenotyping Platforms

The acquisition of trait data can be a time consuming step. To address this issue, many labs have developed robotic phenotyping platforms that allow digital imaging of plant growth. The WIWAM (Weighing, Imaging, and Watering Machine) is equipped to perform controlled watering of 216 plants, along with daily imaging of the shoot tissue, in a greenhouse setting (SKIRYCZ *et al.* 2011). Similarly, the PHENOPSIS platform combines weighing, watering, and imaging of plants with the automatic recording of environmental data (GRANIER *et al.* 2006). In this setup, a camera mounted to a robotic arm enables the study of developmental rates, and when combined with the environmental data, can aid in understanding the interaction between genotype and environmental. One drawback to the aforementioned systems is that each plant in the array is growing within a particular microclimate which may have enough effect on the plant's phenotype to mask the genetic effect of its alleles on the trait of interest. To combat this issue, the automated phenotyping platform Phenoscope utilizes a continuous rotation of plants on a table in a greenhouse, with one position in the array used as an imaging station that allows the monitoring of rosette size and expansion rates (TISNÉ *et al.* 2013). Constant movement of plant samples through the array results in highly reproducible experiments by way of reduced environmental heterogeneity that the plants experience. Although these systems have been optimized for use in Arabidopsis growth studies, other institutions, such as

the EU-SPICY project, have developed phenotyping platforms that can measure plants with much larger vegetative growth structures, such as pepper and tomato, with the use of several cameras mounted in a column (VAN DER HEIJDEN *et al.* 2012; VOORRIPS *et al.* 2010). When a plant is automatically delivered to the SPICY imaging station, these cameras collect images along the height of the stem which can then be merged together, and then image analysis algorithms allow for the phenotyping of traits including plant height and leaf area.

Automated phenotyping platforms are also being developed to study root growth and development. The PlaRoM and Phytomorph projects both use robotics combined with infrared-backlights to produce high quality images of seedling root growth (SUBRAMANIAN *et al.* 2013; YAZDANBAKHSH and FISAHN 2009). In PlaRoM, a robotic arm moves two Petri dishes to the camera-microscope imaging units, while for Phytomorph, the camera mounted to a robotic arm visits each Petri dish in a six by six grid in sequence. Both of these projects have also utilized custom image analysis software to decipher root features from the images produced. Although these systems can provide valuable insight into the genetic and environmental factors controlling seedling growth, growing plants on 2D vertical agar surfaces is not a realistic replacement for a soil environment. In the recent past, studies of RSA in field settings required soil-grown plants to be destructively pulled up, their roots washed, and then the root system imaged with cameras or flatbed scanners in a manually-intensive process (LI *et al.* 2005; ZHU *et al.* 2006; ZHU *et al.* 2005). This method is not ideal in that only one time point per sample can be measured, and the spatial orientation of the roots within the rhizosphere is lost. To begin to address these issues, the GROWSCREEN-Rhizo phenotyping robot was developed to study root

systems in 2D of plants grown in rhizotrons (NAGEL *et al.* 2012). In this system, plants are grown in thin, soil-filled containers and imaged through a clear glass plate in an imaging cabinet travelling on a linear axis. A mechanical swivel arm is used to pull each rhizotron into the cabinet where image capture of the shoot and root system takes place. GROWSCREEN-Rhizo can image 60 rhizotrons per hour, and then a combination of automatic and manual detection of roots via a custom software application provides the measurements of root length, rates of root branching, and the spatial orientation of roots within the environmental. Systems such as these are very useful for determining growth patterns of roots in soil, but because only part of the root system is viewable through the imaging windows, they do not provide a complete portrait of the underground plant structure.

Although 2D imaging systems have proven useful in the studies of RSA, novel methods have also been developed to monitor root system growth in 3D, using cylindrical gellum-gum filled containers. These systems use a camera that takes multiple images from different angles of plant samples allowing recreation of the 3D structure of the root system (IYER-PASCUZZI *et al.* 2010). Grassroots Biotechnology has shown that 19 different root traits of 2 week old rice plants can be analyzed using the RootXpose phenotyping system (INGRAM *et al.* 2012). By monitoring the spatial distribution of roots throughout a clear medium, the large diversity present in different natural accessions can be quantified.

Software

Scientists' ability to determine the genetic basis of plant phenotypes using automated methods depends not only on the rapid acquisition of high quality images, but also on the analysis algorithms used to quantify these traits. Successful interactions between plant biologists and computer scientists well-versed in image analysis techniques are critical to this endeavor. Digital images must be created in a way such that the features of interest are easily extractable from the rest of the data. Additionally, care must be taken to acquire these images in as consistent a manner as possible and named such that the experimental and environmental parameters present during the experiment can be matched to the resulting dataset.

Image analysis algorithms generally start with either a global or adaptive thresholding step to isolate the plant tissue from the background of the image. A histogram of pixel intensities present in the image determines the cutoff for plant versus background (OTSU 1979). Global thresholding is appropriate for when the image background is consistent in terms of pixel intensity, whereas adaptive thresholding looks at only a portion of the image to decide what the best threshold level would be for that section, independent of other areas of the image (SONKA *et al.* 2008). After thresholding, the resulting binary image can be skeletonized using a variety of methods to determine the overall structure of the sample (BLUM 1967; LAM *et al.* 1992; MARAGOS and SCHAFER 1986). Either the binary or skeletonized image can then be analyzed for features such as length and width of segments, number of segments, and overall distribution of the features within the image space (SONKA *et al.* 2008).

Analysis algorithms using methods similar to the above description have been used to measure plant structures such as seeds (HERRIDGE *et al.* 2011; HOSSAIN *et al.* 2010; JOOSEN *et al.* 2012; MOORE *et al.* 2013; TANABATA *et al.* 2012), leaves (ALIMI *et al.* 2013; ZHANG *et al.* 2012), and hypocotyls (COLE *et al.* 2011; SANGSTER *et al.* 2008; WANG *et al.* 2009), and many programs to measure root growth have also been developed. Each program is dependent on a precise imaging protocol in order to accurately extract the features of interest and varies in the amount of user input required for successful completion. Some algorithms, such as SmartRoot and RootTrace, require the user to mark one location on the root surface before analysis of the image can begin (FRENCH *et al.* 2009; LOBET *et al.* 2011; NAEEM *et al.* 2011). The presence of lateral roots, however, can greatly complicate the analysis, especially when they grow across each other on a 2D surface. To surpass this, some programs, such as EZRhizo, allow manual editing by the user to connect or disconnect segments to reach a more accurate picture of the natural growth pattern (ARMENGAUD *et al.* 2009). Users can choose between thresholding algorithms and modify their parameters in GiA Roots, an extensible software package for analyzing root traits of plants grown in minirhizotrons (GALKOVSKYI *et al.* 2012). The creators of this software package have also developed a feature to perform 3D reconstructions of root networks of plants imaged from multiple sides (TOPP *et al.* 2013).

Often, a kinematic view of root growth is desired; these types of algorithms depend on images that capture the texture of the root surface in order to define points of interest that migrate across time. Again, these algorithms come in a variety of levels of automation. For instance, the Kineroot software requires a user to map a large number of points on microscope-

scale images of roots with a high level of texture in order to accurately detect curvature and elongation rates along the root tissue (BASU *et al.* 2007). Alternatively, no user defined points are required for RootFlowRT, which uses optic-flow based techniques to calculate a 2D velocity field based on a stack of images taken at high spatial and temporal resolution (VAN DER WEELE *et al.* 2003). These and other software that calculate relative elemental growth rates (CHAVARRIA-KRAUSER *et al.* 2008; MILLER *et al.* 2009; WALTER *et al.* 2002) enable quantification of the minute changes occurring across cell profiles during root growth.

Data Management

The rapid accumulation of genotype and phenotype information enabled by high-throughput phenotyping systems requires consideration of best practices for recording, storage, and sharing of the data. Raw images could be made available for analysis by other labs with different algorithms, with the potential of quantifying new features within those images. Very little raw data from published QTL studies are available for download and verification; one estimate is as few as 1% of QTL papers have the raw data presented for download (ZAMIR 2013). Clearly, there is a need for common experimental databases with smart searching and sorting features that will enable sharing and exploration of genotype-phenotype relationships. Some websites, including the Nottingham Arabidopsis Stock Centre (NASC) and the RIKEN Arabidopsis Phenome Information Database host images of mutant plants that show phenotypes different from that of a wild type plant (KUROMORI *et al.* 2006), however their focus is currently on the above ground plant structures. Dryad (http://datadryad.org/) and QTL Archive

(http://qtlarchive.org) serve as general purpose repositories of QTL data. QTL Archive is mainly used for rodent data, while Dryad hosts data from a variety of organisms. Dryad also is integrated with numerous journals to facilitate the process of peer review while also allowing the public sharing of the data. The Phenom-Networks database (http://phnserver.phenome-networks.com) stores raw phenotypic data from QTL experiments performed with a variety of model organisms, including data from 355 traits measured in 45 independent QTL experiments using tomato introgression lines. Likewise, the website http://www.genenetwork.org hosts published QTL datasets from several major model species, including Arabidopsis, and also provides online QTL mapping tools.

Often, databases are designed as complements to a phenotyping platform, such as with PHENOPSIS DB, a freely available database designed to store, retrieve, and share the phenotype and metadata collected with the PHENOPSIS platform (FABRE *et al.* 2011). The need for flexibility in the types of raw and metadata is addressed with Phenomics Ontology Driven Data Management (PODD), a data management architecture that using ontology-based domain modeling for scientific data management (LI *et al.* 2013). Similarly, project-specific databases can be designed with Phenotyper, a free data management system that was created to address technical challenges present in a multi-site phenotyping experiment (BILLIAU *et al.* 2012). This project realized the need for central data storage and web-enabled searching for scientists from 11 different groups. By using controlled ontologies in describing experimental results, data from these and other sources have the ability to be fully integrated for better information gathering.

Although one major problem for researchers interested in phenotyping is storage and access of data discussed in peer-reviewed journals, another major issue is the ability to capture all the relevant experimental data in the first place. For instance, the genotype, age, growth substrate, and replicate number are all obvious variables for an experiment, but the season, time of day, humidity, and temperature, among others, are also important aspects of the metadata that can influence the interpretation of results, and therefore should be measured, recorded, and shared. The Phenome software application utilizes a Personal Digital Assistant (PDA) to record data from large scale phenotyping experiments (VANKADAVATH *et al.* 2009). Using a built-in barcode reader, users can scan the label on each sample and record information on user-selected phenotypes using a touch-screen interface. Digital storage of phenotype values combined with automatically logging environmental data will be crucial to the full understanding of a plant's phenotypic plasticity in response to differing environmental conditions.

Conclusions

The number of mapping populations available for genetic studies is vast compared to our technical abilities to quantify their diversity. Destructive and manual methods are quickly being replaced by digital phenotyping capabilities and custom software to extract high quality trait data from the images produced. The need for information storage and retrieval platforms is now an essential factor governing our ability to interpret the way a plant's genetics interact with the environment, as little time will be needed to accumulate a large amount of raw and

experimental metadata. Ideally, all data from published experiments would be available through an interactive online database for others to mine for accuracy and new phenotypes. By structuring our information gathering efforts in a consistent manner, we gain the ability to examine phenotypic trends across species boundaries to uncover functions of previously uncharacterized genes. Harnessing the natural variation of key loci will be imperative in enabling a new revolution in crop improvement necessary for continued global food security.

Preface to Chapter 2

Chapter 1 discussed the important elements of a successful phenotyping platform, including the ability to automate a high-throughput system of imaging and feature extraction, create appropriate data management structures, and share genetic and phenotype data with the community. In Chapter 2, I describe a phenotyping platform for measuring seed size descriptors using flatbed scanners. By quantifying the parameters of seed size in the model plant Arabidopsis, we were able to identify key loci that contribute to the variation in these traits. The methods and data presented here enhanced our understanding of how seed size is determined and will serve as a useful model for future studies of seed phenotypes.

Chapter 2:

Mapping quantitative trait loci affecting *Arabidopsis thaliana* seed morphology features extracted computationally from images

Abstract

Seeds are studied to understand dispersal and establishment of the next generation, as units of agricultural yield, and for other important reasons. Thus, elucidating the genetic architecture of seed size and shape traits will benefit basic and applied plant biology research. This study sought quantitative trait loci (QTL) controlling the size and shape of *Arabidopsis thaliana* seeds by computational analysis of seed phenotypes in recombinant inbred lines derived from the small-seeded Landsberg *erecta* x large-seeded Cape Verde Islands accessions. On the order of 10^3 seeds from each recombinant inbred line were automatically measured with flatbed photo scanners and custom image analysis software. The eight significant QTL affecting seed area explained 63% of the variation, and overlapped with five of the six major-axis (length) QTL and three of the five minor-axis (width) QTL, which accounted for 57 and 38% of the variation in those traits, respectively. Because the Arabidopsis seed is exalbuminous, lacking an endosperm at maturity, the results are relatable to embryo length and width. The Cvi alleles generally had a positive effect of 2.6–4.0%. Analysis of variance showed heritability of the three traits ranged between 60% and 73%. Repeating the experiment with 2.2 million seeds from a separate harvest of the RIL population and approximately 0.5 million seeds from

92 near isogenic lines confirmed the above results. Structured for download are files containing phenotype measurements, all sets of seed images, and the seed trait measuring tool.

Introduction

The seed disperses, protects, and sustains the beginning stage of the next generation in plants that produce them. Post-germination survival and subsequent development of the next generation depends on seed parameters such as size (KRANNITZ *et al.* 1991; MANNING *et al.* 2009). Also, the seed is the primary product of many agricultural crops, and biotechnologists are endeavoring to engineer its chemical composition and properties (ABELSON and HINES 1999; UFAZ and GALILI 2008). Thus, an important goal in basic and applied plant biology is to elucidate the genetic elements responsible for controlling seed size and shape. Quantitative trait locus (QTL) analysis based on phenotypic data from many plants harboring different and known combinations of two distinct parental DNA types is a proven approach to achieving this goal (ALONSO-BLANCO *et al.* 2009). Many excellent populations of such recombinant inbred lines (RILs) have been created among divergent accessions of *Arabidopsis thaliana* because this species' short generation time, natural self-pollination trait, and numerous diverse accessions make it very conducive to QTL analysis (KOVER and MOTT 2012; MEYEROWITZ 2001). For example, the collection of 162 Arabidopsis RILs derived from a cross between the Landsberg *erecta* (L*er*) and Cape Verde Islands (Cvi) accessions supports mapping of QTL to genomic intervals smaller than 1 cM (ALONSO-BLANCO *et al.* 1998). Thus, the germplasm and genotype information are strengths of Arabidopsis QTL studies. What typically limits the quality of such studies is the

quality of the phenotype dataset. Traits can be difficult to measure repeatedly and precisely in the many members of the RIL population. In the case of seeds, this is especially true because each elliptical Arabidopsis seed is less than a millimeter long. A pioneering seed size QTL study was previously performed with the Cvi x L*er* RIL population (ALONSO-BLANCO *et al.* 1999) in which seed size was measured by microscope-aided inspection. The limitations in throughput and precision inherent in such a manual technique can be alleviated by using a flatbed photo scanner to acquire digital images containing many seeds in a field of view, and image processing techniques to quantify features such as the projected area of each separate seed (HERRIDGE *et al.* 2011; SCHNEIDER *et al.* 2012). The project reported here represents an extension of the growing trend to use computational methodologies for precision phenotype measurements and the especially robust genotype-to-phenotype mapping that this approach enables. In addition to reporting the genetic architecture of the morphology of Arabidopsis seeds and therefore of the embryo plant within, the present report makes available to the community a comprehensive image set (the raw data), the quantified morphological features of many thousands of seeds from the Cvi x L*er* population (the processed trait data), and a software tool for creating the latter from the former.

Materials and Methods

Plant material

This study used the 162 genotyped RILs derived from the L*er* and Cvi inbred parents (ALONSO-BLANCO *et al.* 1998). The seeds used in the population referred to in this study as RIL1

were kindly provided by Dr. Patrick Masson from the University of Wisconsin-Madison, Madison, WI and were propagated in parallel inside a 22° growth chamber as described in Vaughn and Masson (2011). Data from the population referred to in this study as RIL2 were collected from a second harvest of seeds produced by six replicate plants for each RIL grown in a randomized pattern in an air-conditioned greenhouse environment at the University of Wisconsin-Madison Biotron. Temperature was maintained at 23°C during the 16 h day and 21°C during the night. The beds were watered with 0.25X Hoagland's solution twice a week for the first month, followed by once a week for the remaining growth period. At appearance of the first flower, each pot was enclosed using ArabiSifters (SNS-02, Lehle Seeds, Round Rock TX) to prevent cross-pollination. At maturity, each plant was dried in a 90° room and its seeds were sifted through a coarse mesh to remove plant debris before being placed in a plastic tube for storage. Additionally, a set of 92 NILs was obtained through ABRC. Seeds from each NIL were produced and harvested as described for the method of RIL2 generation.

Seed trait measurements

Approximately 1000 seeds from each mother plant were sprinkled onto a square Petri dish, and scanned using an Epson Perfection 4990 Photo series scanner to obtain an 8 bit grayscale image. RIL populations were scanned at 3200 dots per inch (dpi), and the NIL population was scanned at 4800 dpi. A custom computer program was developed to detect seeds in the images thus produced and measured the seed area, seed length (major axis), and seed width (minor axis) for each detected seed. To summarize, the tasks performed by the algorithm for extracting and measuring the seeds present in an image are:

1) obtain and store a threshold value that segments the grayscale image into foreground (seeds) and background;

2) identify each potential seed in the binarized image and determine its area, minor axis, and major axis; and

3) filter results to reject image components not corresponding to individual seeds.

Task 1 is a multistep process because variation in the background values and the number of seeds present produces variation in the pixel intensity histograms used to select the most effective threshold value by a standard method (OTSU 1979). No single threshold value is optimal for all images and if an image contains very few seeds, and therefore a low number of black pixels, the Otsu method may not produce a useful result. To obtain a threshold value suitable for all images in a batch, the optimal threshold value for each individual image is determined. These individual values are averaged and the mean threshold is used to binarize all images in the batch. Task 2 identifies each component of the image consisting of 8-connected black pixels, the potential seeds, and computes the area, major axis, and minor axis of each of these 8-connected components. The results are saved to disk. Task 3 determines which of the identified and measured components has the characteristics of an ellipsoidal seed. From the major and minor axes determined for each 8-connected component, the area is calculated assuming the object is ellipsoidal. This modeled area is compared with the measured area. If the two agree, the component passes the first stage of filtering, which effectively removes instances of two touching seeds and irregularly-shaped debris. The second

filtering step removes components that pass the ellipse test but which have a major axis more than 2.5 times longer than the minor axis, and therefore are not seeds. A scratch in the Petri plate is an example of an artifact that heuristic was implemented to remove. The area, major axis, and minor axis derived from image components after filtering to remove non-seeds were the traits subjected to statistical genetic analyses.

The algorithm, written in the Matlab® language, and operating instructions are presented for download at http://phytomorph.wisc.edu/G3. Also presented there for download are all the raw images and the comma separated value files containing the filtered results obtained from each image.

QTL analysis

After image analysis, the mean seed area, major axis, and minor axis were computed for each set of progeny. For the RIL2 and NIL populations, the final phenotype value for each line is the average of all the replicates. Genotype information from 234 markers in the L*er*/Cvi RIL map and 102 markers from the L*er*/Cvi NIL map was collected from previously published work (ALONSO-BLANCO *et al.* 1998; KEURENTJES *et al.* 2007). The qtl library (BROMAN *et al.* 2003) within the R statistical software (www.r-project.org/) was used to search for and characterize significant loci linked to the markers using multiple-QTL mapping methods. Two hundred and fifty-six rounds of imputation (SEN and CHURCHILL 2001) were performed using pseudomarkers at 1 cM intervals. A genotyping error rate of 0.001 was assumed, and the Kosambi map function (KOSAMBI 1944) was used to estimate genetic distances. Significance thresholds (α=0.05) were calculated by a permutation test (CHURCHILL and DOERGE 1994) of a two-dimensional, two-QTL

scan analyzed by Haley-Knott regression (HALEY and KNOTT 1992) using 25,000 permutation replicates. Applying significance thresholds calculated in this manner produced essentially the same results as those obtained by the imputation method. The best QTL model was selected using the stepwise QTL analysis of Manichaikul *et al.* (2009) with forward model selection proceeding up to 10 QTL followed by backward elimination. This approach seeks as many true QTL as possible, minimizes the inclusion of extraneous loci, but is permissive of extraneous interactions joining the model as the number of potentially interacting QTL increases.

The quality of each potential model was evaluated using a penalized logarithm of odds (LOD) score that balances model fit and model complexity by subtracting a penalty derived from the permutation tests for each additional QTL or QTL:QTL interaction present in the model. The LOD penalties were calculated for main effect QTL and epistatic interactions on the basis of the thresholds derived from the scantwo permutation tests. The main, heavy, and light interaction penalties as well as the 5% significance thresholds are presented in accordance with Broman and Sen (2009) in the appropriate table legend. The models generated by the stepwise QTL method of Manichaikul *et al.* (2009) include only QTL and interactions deemed to be significant using the permutation thresholds previously calculated. The chosen model is the one which has the highest penalized LOD score among all models evaluated. Positions of QTL in the final model were refined then fit to the phenotypic data to provide estimates of each QTL's effect and LOD score from the fit of the full model. A 1.5 LOD support interval was used to determine the confidence intervals for each locus (MANICHAIKUL *et al.* 2006).

Results

Figure 1 shows scanning electron micrographs of two representative seeds from each of the parental accessions used to create the RILs studied here. The images show a typical size difference between Cvi and L*er* seeds and convey how the embryo folded within determines the size and shape of the exalbuminous seed (ESAU 1953). It is clear from these images that the length (major axis) reflects the length of the hypocotyl-root axis and the width (minor axis) is essentially the sum of the cotyledon and hypocotyl widths. Thus, a high-resolution quantification of Arabidopsis seed morphology will relate to embryo structure to a considerable extent.

The scanning electron microscope captures exquisite structural details of the Arabidopsis seed but the imaging procedures cannot be performed with sufficient throughput and automation to support a statistical study of the QTL affecting seed morphology. An appropriate combination of resolution and throughput was achieved with a flatbed photo scanner set to acquire images of seeds scattered in a clear dish at 3200 dpi. Figure 2 shows that the seeds in each of the resulting images were resolved well enough that a computer algorithm could be coded to determine the pixels comprising the contours of each individual seed and reject instances of two or more touching seeds (Figure 2A *inset*). The algorithm automatically returned the area, major axis length, and minor axis length of each successfully segmented, individual seed. On average, 1600 seeds were measured per image. The algorithm in the form of two Matlab® code files is presented for download at http://phytomorph.wisc.edu/G3.

Figure 1: Scanning electron micrographs of two seeds from the Ler and two seeds of the Cvi accession. c, cotyledon; h, hypocotyl; r, embryonic root or radical. Scale bar = 300 μm.

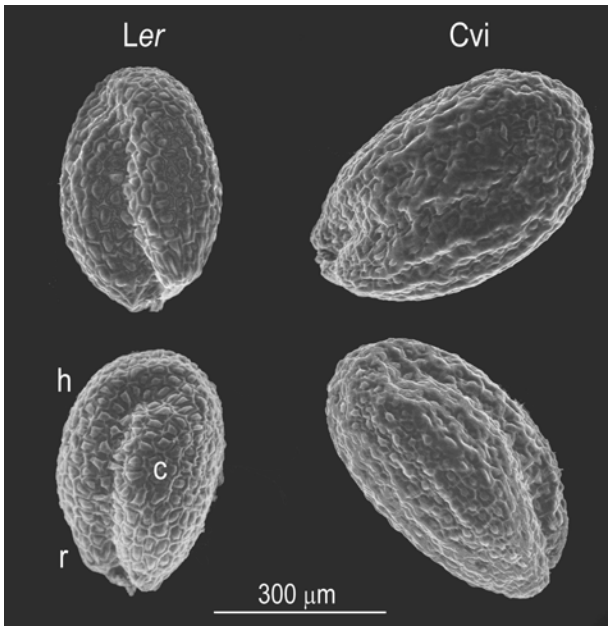
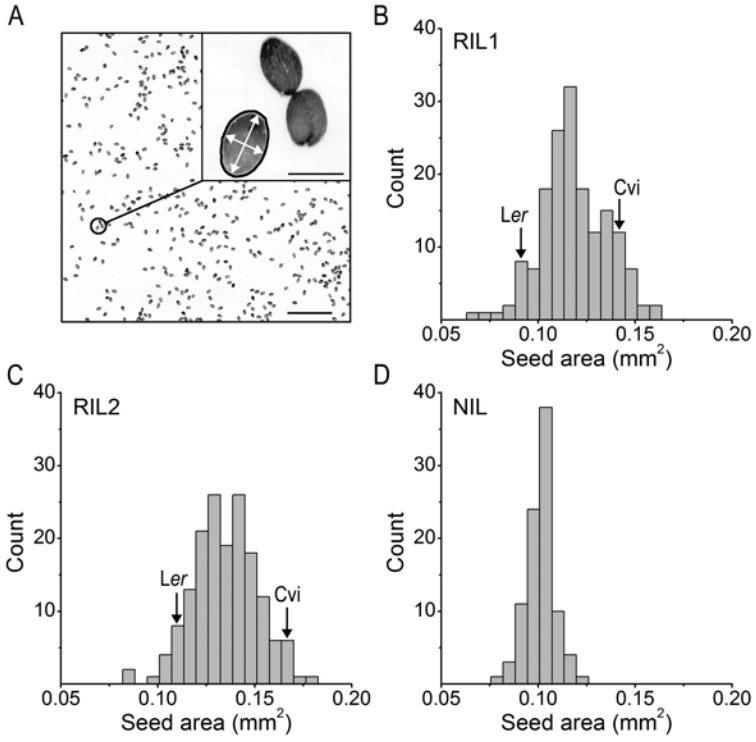


Figure 2: Arabidopsis seed area and shape measured by automatic image processing. (A) A sample of a typical image of a field of Arabidopsis seeds acquired with a flatbed photo scanner. Scale bar = 5 mm. (Inset) An expansion of the field showing a cluster of three seeds. A custom algorithm determines the boundary, or contour (black line), of all objects in the image that can be separated (segmented) from the background and rejects any not having the morphological properties of a single seed such as the two touching seeds shown, or a piece of debris. The white arrows indicate the major and minor axes of a successfully segmented individual seed. Scale bar = 0.5 mm. (B–D), Frequency distribution of seed area in two independent generations of the Cvi x L*er* RIL population, RIL1 and RIL2, respectively (B, C) and a Cvi x L*er* NIL population (D). The 95% confidence interval of the parental means lies within the specified bins.



The Cvi x L*er* population consists of 162 distinct RILs (ALONSO-BLANCO *et al.* 1998). Aliquots of seeds from all but one (#157) of the RILs were separately scanned and the images processed to produce a dataset containing the average area, major axis length, and minor axis length for each RIL. This dataset is referred to hereafter as RIL1. Figure 2B shows the frequency distribution of the area trait within RIL1. The Cvi and L*er* parental means are indicated to show the extent of transgressive segregation in the population. A separate harvest of this population of RILs was generated and aliquots of their seeds scanned and measured to create the RIL2 dataset. Figure 2C shows the frequency of the seed area trait within RIL2. Near isogenic lines (NILs) are frequently helpful in detecting small-effect QTL. Therefore, 92 NILs each containing one small genomic region of Cvi introgressed into L*er* (KEURENTJES *et al.* 2007) were raised to produce seeds. These seeds were scanned as before, except that scanning was at 4800 dpi. Figure 2D shows the frequency of the seed area trait within the resulting NIL dataset.

Analysis of variance was performed to determine the heritability (H^2) of the traits in each of the datasets (Table 1) (FALCONER and MACKAY 1996). For RIL1 and RIL2, H^2 ranged from a low of 52% (RIL2 minor axis) to a high of 73% (RIL1 area). Heritability in NIL was much lower due to the limited genetic variation between introgression lines.

The RIL and NIL populations used here had previously been genotyped with 234 and 102 markers (ALONSO-BLANCO *et al.* 1998; KEURENTJES *et al.* 2007), respectively, permitting QTL analysis of the three datasets using multiple interval mapping. Statistical significance of the QTL models was based on 25,000 permutations of the genotype against each phenotype, for

Table 1: Heritability for seed shape traits estimated from analysis of variance. MS_M, mean square model; MS_E, mean square error; df_M, degrees of freedom of model; df_E, degrees of freedom of error, V_A, additive genetic variance; V_E, environmental variance; H^2, heritability.

Trait	MS_M	MS_E	df_M	df_E	V_A	V_E	H^2
				RIL1			
Area	0.266	2.28×10^{-4}	160	1.64×10^{5}	2.60×10^{-4}	9.78×10^{-5}	0.727
Major Axis	1.54	1.52×10^{-3}	160	1.64×10^{5}	1.50×10^{-3}	7.68×10^{-4}	0.662
Minor Axis	0.421	4.79×10^{-4}	160	1.64×10^{5}	4.12×10^{-4}	2.73×10^{-4}	0.601
				RIL2			
Area	3.61	2.84×10^{-4}	161	2.18×10^{6}	2.69×10^{-4}	1.50×10^{-4}	0.642
Major Axis	17.3	1.57×10^{-3}	161	2.18×10^{6}	1.29×10^{-3}	9.25×10^{-4}	0.582
Minor Axis	5.93	6.26×10^{-4}	161	2.18×10^{6}	4.41×10^{-4}	4.05×10^{-4}	0.521
				NIL			
Area	0.265	2.24×10^{-4}	91	4.77×10^{5}	5.10×10^{-5}	1.98×10^{-4}	0.205
Major Axis	2.28	1.79×10^{-3}	91	4.77×10^{5}	4.40×10^{-4}	1.57×10^{-3}	0.219
Minor Axis	0.403	6.02×10^{-4}	91	4.77×10^{5}	7.76×10^{-5}	5.63×10^{-4}	0.121

each dataset (CHURCHILL and DOERGE 1994). LOD score profiles for each of the traits in each of the datasets are shown in Figure 3. Asterisks denote significant QTL. Allele effects associated with each QTL genotype are shown in Figure 4. The 1.5 LOD support intervals associated with each QTL are shown in Figure 5. The genomic position, associated marker, additive allele effect, and percentage of explained variance for each QTL are presented for the area trait (Table 2), major axis (Table 3), and minor axis (Table 4). Although the qtl library in R is not yet optimized for NIL populations, using an identical protocol as with the RIL population and interpreting the NIL results with caution allowed for ease of comparison between the three seed populations.

For the area trait in RIL1, eight QTL were found to be significant ($P < 0.001$), including three on chromosome 1, two on chromosome 3, one on chromosome 4, and two on chromosome 5. The method of QTL identification used here is also capable of identifying interactions between loci, i.e. epistatic relationships in which the effect on a trait of a genotype at one position influences the genotype effect at a second locus. Evidence of epistatic interactions between two pairs of loci were found in RIL1 for the area trait. They are indicated in Table 2 by a colon separating the labels of the two interacting loci and characterized in Figure 6A and B. For example, a QTL at 19.6 cM on chromosome 1 interacted with a QTL at 4.0 cM on chromosome 3, or 1@19.6:3@4.0. For the major axis trait in RIL1, six QTL were supported ($P < 0.001$), with one instance of epistasis detected (Table 3). QTL analysis of the minor axis trait in RIL1 revealed five significant loci ($P < 0.001$) with no evidence of epistasis.

For RIL2, QTL analysis of seed area resulted in 9 significant loci ($P < 0.001$), including three on chromosome 1, one on chromosome 2, two on chromosome 3, two on chromosome 4,

Figure 3: LOD profiles of seed traits statistically modeled as a function of genotype. (A–C), Seed area QTL determined by multiple-interval mapping using the two independent RIL datasets (A, B) and the NIL dataset (C). (D–F), Major axis QTL of the RIL datasets (D, E) and the NIL dataset (F). (G–I), Minor axis QTL of the RIL datasets (G, H), and the NIL dataset (I). Asterisks denote the position with the highest LOD score for each locus. Vertical dotted lines are used to separate the five chromosomes. Horizontal, dotted lines indicate the significance threshold.

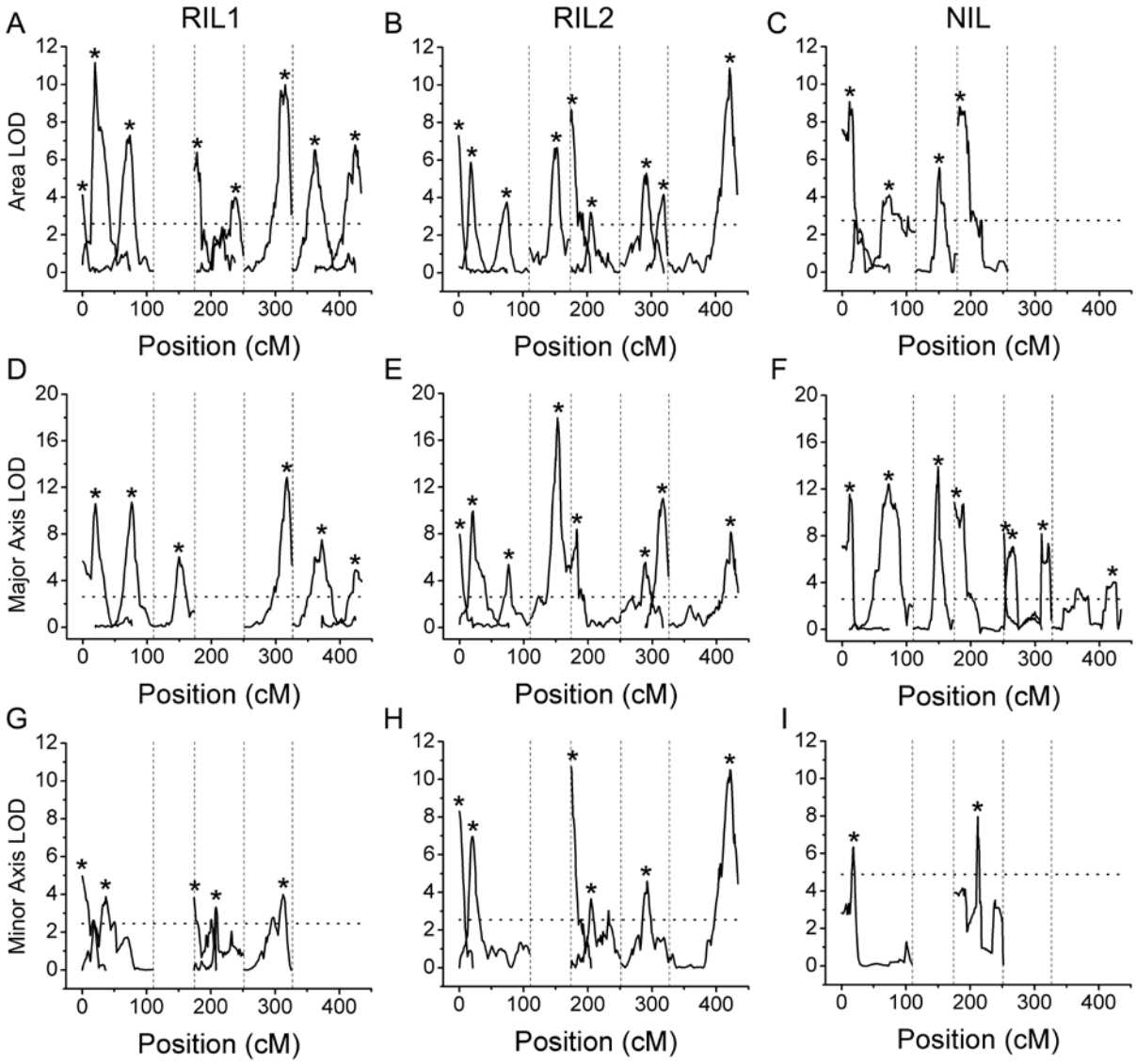


Figure 4: Estimated effects on the seed area trait of each marked position of the genome determined from the indicated dataset. Positive values indicate that substitution of a Cvi allele increases the trait; negative values indicate a L*er* allele at that position increases the trait in the indicated dataset. (A–C), Allele effects on seed area determined from the two independent RIL datasets (A, B) and the NIL dataset (C). (D–F), Allele effects on major axis QTL of the RIL datasets (D, E) and the NIL dataset (F). (G–I), Allele effects on minor axis QTL of the RIL datasets (G, H) and the NIL dataset (I).

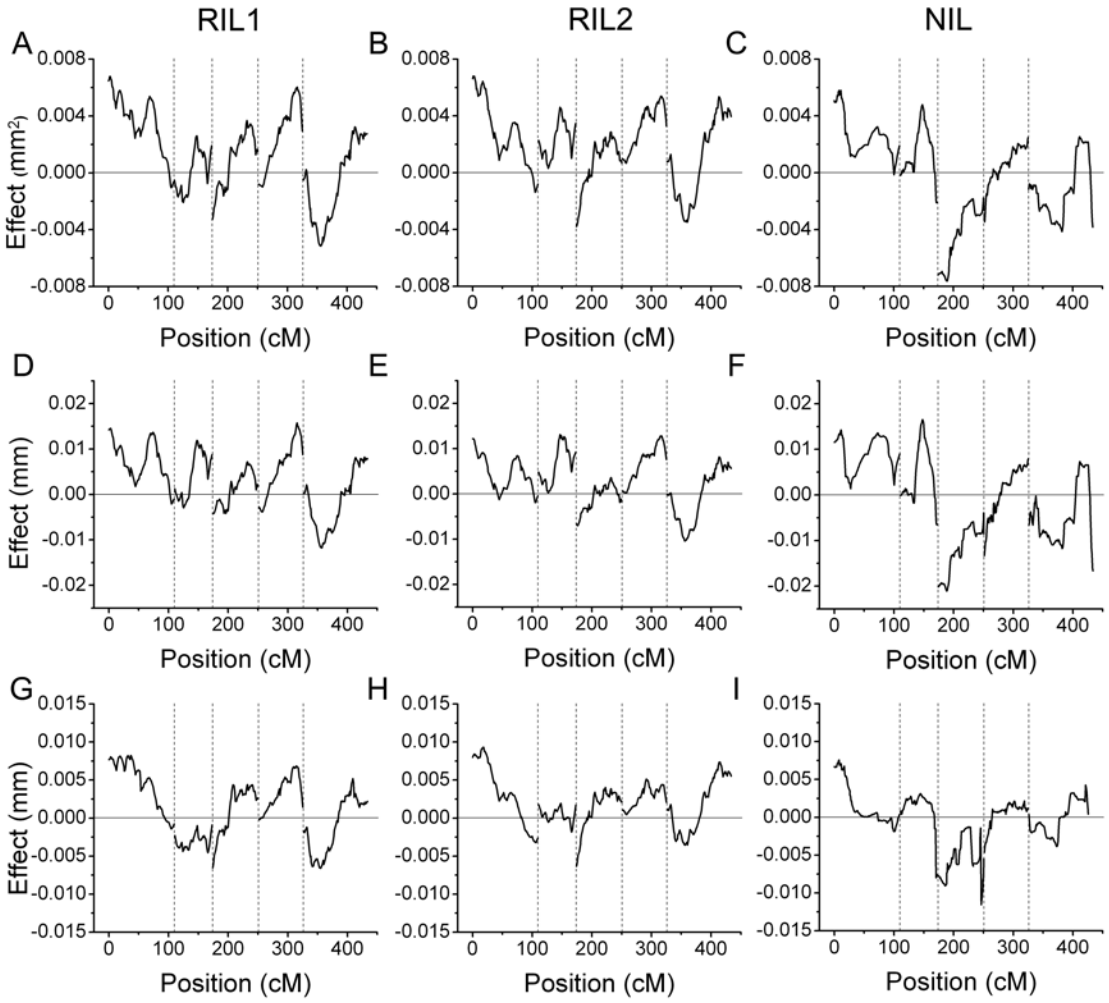


Figure 5: LOD support intervals of significant QTL identified by multiple-interval mapping for each indicated dataset. Regions were determined using a 1.5-LOD support interval, where the QTL is in the region in which the LOD score is within 1.5 of its maximum.

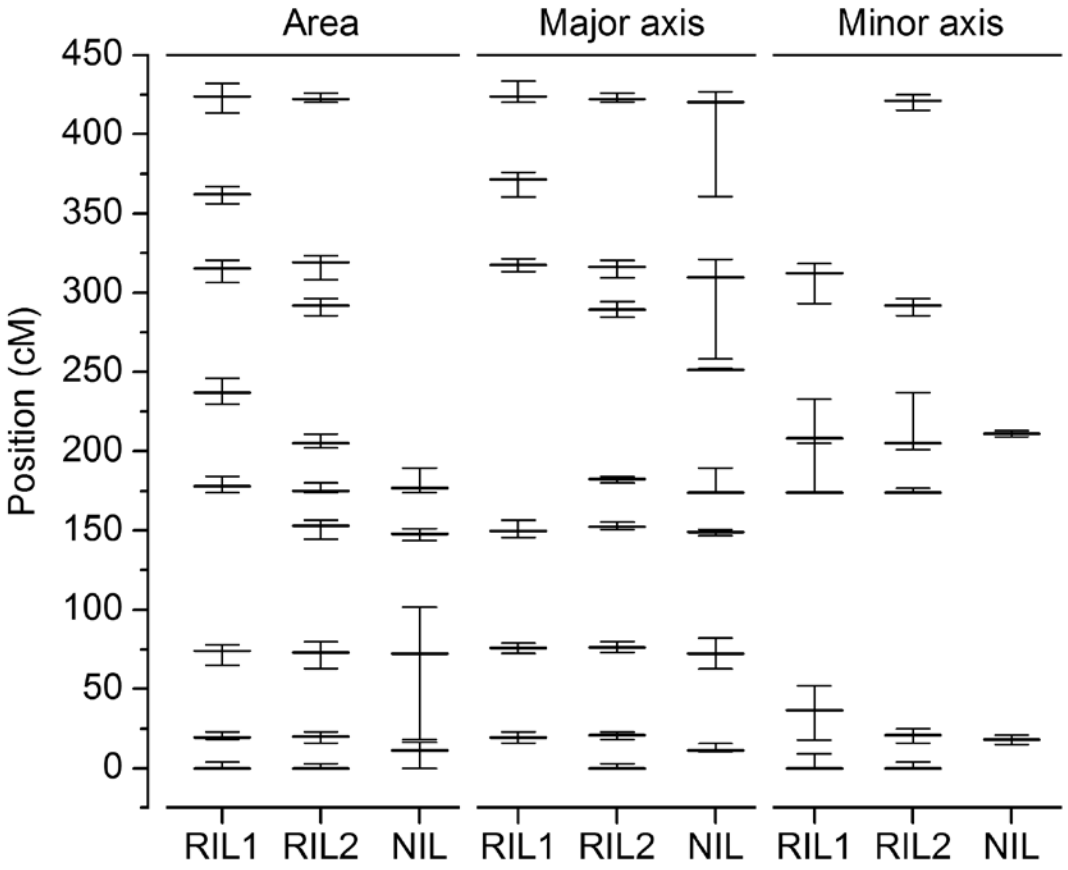


Table 2: QTL affecting seed area. The genomic position, associated marker or pseudomarker, estimated additive effect of substitution of a Cvi allele at the indicated locus, and percentage of explained variance for each QTL present in the selected model with the highest LOD score. For RIL1, (T_f, T_{fv1}, T_i, T_a, T_{av1}) = (5.51, 4.12, 3.40, 4.36, 2.63) and (T_m, T_i^H, T_i^L) = (2.57, 3.40, 1.55). For RIL2, (T_f, T_{fv1}, T_i, T_a, T_{av1}) = (5.50, 4.12, 3.42, 4.34, 2.58) and (T_m, T_i^H, T_i^L) = (2.60, 3.42, 1.52). For NIL, (T_f, T_{fv1}, T_i, T_a, T_{av1}) = (5.04, 3.51, 2.59, 4.55, 2.57) and (T_m, T_i^H, T_i^L) = (2.74, 2.49, 0.76).

Position	Associated marker	Additive effect (mm^2)	Variance explained (%)	LOD
RIL1				
1@0.0	PVV4	0.0043	4.6	4.1
1@19.6	GD.86L	0.0067	13.9	11.2
1@74.0	c1.loc74	0.0052	8.6	7.3
3@4.0	c3.loc4	-0.0041	7.4	6.4
3@63.0	c3.loc63	0.0036	4.5	4.0
4@64.0	c4.loc64	0.0060	12.2	10.0
5@36.0	c5.loc36	-0.0039	7.6	6.5
5@98.0	c5.loc98	0.0046	7.9	6.8
1@19.6:3@4.0	GD.86L: c3.loc4	0.0043	5.3	4.7
5@36.0:5@98.0	c5.loc36: c5.loc98	0.0020	1.5	1.4
RIL2				
1@0.0	PVV4	0.0054	8.5	7.6
1@20.0	c1.loc20	0.0048	6.5	5.8
1@73.0	c1.loc73	0.0033	3.9	3.6

2@42.7	Erecta	0.0043	7.0	6.3
3@1.0	c3.loc1	-0.0055	10.1	8.7
3@31.0	AD.92L	0.0033	3.6	3.4
4@40.6	DF.108L-Col	0.0041	6.3	5.7
4@67.8	GB.750C	0.0034	3.9	3.6
5@96.0	c5.loc96	0.0060	12.9	10.9
NIL				
1@11.2	m6	0.0077	21.0	9.1
1@72.2	c1.loc72	0.0051	8.4	4.1
2@37.3	m42	0.0051	11.8	5.6
3@2.9	c3.loc3	-0.0059	20.3	8.8
1@11.2: 1@72.2	m6: c1.loc72	0.0026	2.1	1.1

Table 3: QTL affecting seed length (major axis). The genomic position, associated marker or pseudomarker, estimated additive effect of substitution of a Cvi allele at this locus, and percentage of explained variance for each QTL present in the selected model with the highest LOD score. For RIL1, (T_f, T_{fv1}, T_i, T_a, T_{av1}) = (5.51, 4.12, 3.45, 4.33, 2.55) and (T_m, T_i^H, T_i^L) = (2.59, 3.45, 1.53). For RIL2, (T_f, T_{fv1}, T_i, T_a, T_{av1}) = (5.49, 4.09, 3.41, 4.34, 2.57) and (T_m, T_i^H, T_i^L) = (2.58, 3.41, 1.51). For NIL, (T_f, T_{fv1}, T_i, T_a, T_{av1}) = (4.65, 3.10, 2.16, 4.23, 2.20) and (T_m, T_i^H, T_i^L) = (2.56, 2.16, 0.54).

Position	Associated marker	Additive effect (mm)	Variance explained (%)	LOD
RIL1				
1@19.6	GD.86L	0.011	15.3	10.6
1@76.0	c1.loc76	0.018	15.5	10.7
2@39.0	c2.loc39	0.012	8.1	6.0
4@66.0	c4.loc66	0.018	19.2	12.9
5@45.5	HH.480C	-0.0066	10.4	7.5
5@98.0	c5.loc98	0.011	6.5	4.9
1@19.6: 5@45.5	GD.86L: HH.480C	-0.011	6.6	5.0
RIL2				
1@0.0	PVV4	0.011	8.1	8.0
1@21.0	c1.loc21	0.013	10.5	10.0
1@76.3	GD.160C	0.0082	5.3	5.4
2@42.0	c2.loc42	0.016	21.3	17.9
3@8.3	CH.322C	-0.011	8.6	8.4

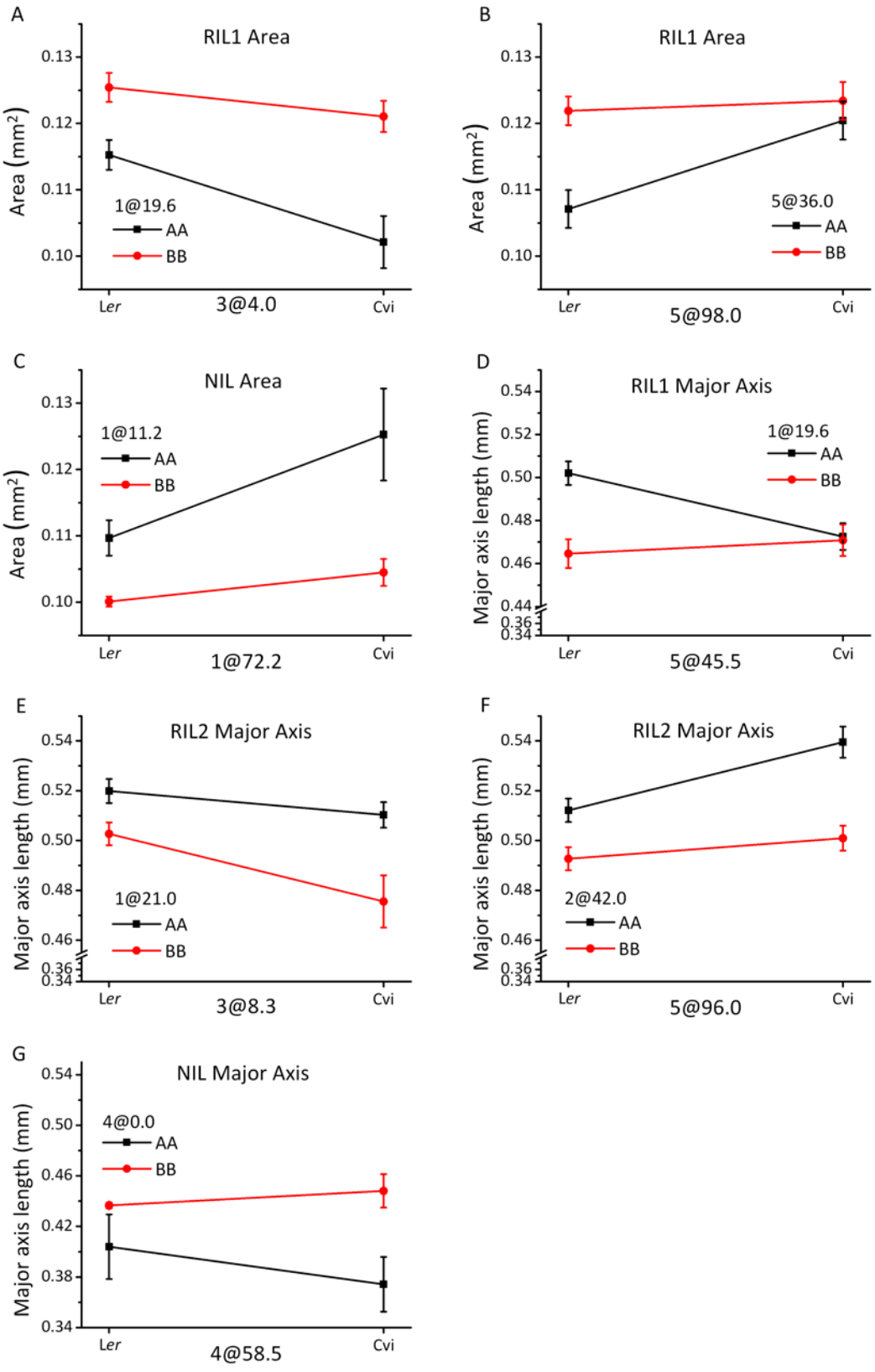
4@38.0	c4.loc38	0.0084	5.5	5.6
4@65.0	c4.loc65	0.013	11.8	11.0
5@96.0	c5.loc96	0.0099	8.3	8.1
1@21.0: 3@8.3	c1.loc21: CH.322C	0.0094	5.2	5.3
2@42.0: 5@96.0	c2.loc42: c5.loc96	0.0047	1.6	1.7
NIL				
1@11.3	c1.loc11	0.015	16.2	11.6
1@72.2	c1.loc72	0.014	17.8	12.4
2@38.5	c2.loc39	0.019	20.8	13.9
3@0.0	m53	-0.015	14.9	10.9
4@0.0	m69	-0.015	10.4	8.2
4@58.5	c4.loc59	0.023	10.4	8.2
5@95.2	c5.loc95	0.0095	4.6	4.0
4@0.0: 4@58.5	m69: c4.loc59	0.012	3.6	3.2

Table 4: QTL affecting seed width (minor axis). The genomic position, associated marker or pseudomarker, estimated additive effect of substitution of a Cvi allele at this locus, and percentage of explained variance for each QTL present in the selected model with the highest LOD score. For RIL1, (T_f, T_{fv1}, T_i, T_a, T_{av1}) = (5.83, 4.58, 3.56, 4.49, 2.97) and (T_m, T_i^H, T_i^L) = (2.55, 3.56, 2.03). For RIL2, (T_f, T_{fv1}, T_i, T_a, T_{av1}) = (5.75, 4.50, 3.56, 4.47, 2.95) and (T_m, T_i^H, T_i^L) = (2.55, 3.56, 1.95). For NIL, (T_f, T_{fv1}, T_i, T_a, T_{av1}) = (7.80, 5.96, 4.63, 6.73, 4.89) and (T_m, T_i^H, T_i^L) = (4.91, 4.63, 1.06).

Position	Associated marker	Additive effect (mm)	Variance explained (%)	LOD
RIL1				
1@0.0	PVV4	0.0077	9.4	5.0
1@36.6	AD.106L-Col	0.0068	7.3	3.9
3@0.0	DF.77C	-0.0068	7.1	3.8
3@34.0	c3.loc34	0.0063	6.1	3.3
4@61.0	c4.loc61	0.0068	7.5	4.0
RIL2				
1@0.0	PVV4	0.0077	10.9	8.3
1@21.0	c1.loc21	0.0072	9.0	7.0
3@0.0	DF.77C	-0.0082	14.5	10.7
3@31.0	AD.92L	0.0048	4.5	3.7
4@40.6	DF.108L-Col	0.0051	5.7	4.6

5@95.0	c5.loc95	0.0082	14.2	10.5
NIL				
1@18.3	c1.loc18	0.0091	22.2	6.3
3@37.1	c3.loc37	-0.0085	29.2	8.0

Figure 6: Genotype-phenotype plots for epistatic QTL pairs in the L*er* x Cvi population. Plotted points indicate two-locus genotype means ± SE for each pair of loci from the RIL1 area trait (A, B) NIL area trait (C), RIL1 major axis trait (D), RIL2 major axis trait (E, F) and NIL major axis trait (G).


A
RIL1 Area
Area (mm²)
0.13
0.12
0.11
0.10
1@19.6
AA
BB
Ler
3@4.0
Cvi
B
RIL1 Area
Area (mm²)
0.13
0.12
0.11
0.10
5@36.0
AA
BB
Ler
5@98.0
Cvi
C
NIL Area
Area (mm²)
0.13
0.12
0.11
0.10
1@11.2
AA
BB
Ler
1@72.2
Cvi
D
RIL1 Major Axis
Major axis length (mm)
0.54
0.52
0.50
0.48
0.46
0.44
0.36
0.34
1@19.6
AA
BB
Ler
5@45.5
Cvi
E
RIL2 Major Axis
Major axis length (mm)
0.54
0.52
0.50
0.48
0.46
0.36
0.34
1@21.0
AA
BB
Ler
3@8.3
Cvi
F
RIL2 Major Axis
Major axis length (mm)
0.54
0.52
0.50
0.48
0.46
0.36
0.34
2@42.0
AA
BB
Ler
5@96.0
Cvi
G
NIL Major Axis
Major axis length (mm)
0.54
0.50
0.46
0.42
0.38
0.34
4@0.0
AA
BB
Ler
4@58.5
Cvi

and one on chromosome 5. No evidence of epistasis was found. For the major axis trait in RIL2, eight QTL were chosen ($P < 0.001$), with evidence of epistasis between two pairs of loci (Figure 6D and E). For the minor axis trait, six QTL were identified ($P < 0.001$) with no evidence of epistasis.

For NIL, four significant loci were identified as contributing to the variation in seed area ($P < 0.001$), with two loci on chromosome 1, one locus on chromosome 2, and one locus on chromosome 3. Evidence of epistasis was found between the two loci on chromosome 1 (Table 2 and Figure 6C). Seven QTL were found to contribute to the variation in major axis in the NIL population ($P < 0.001$), with epistasis likely between the two loci on chromosome 4 (Table 3 and Figure 6G). For the minor axis trait, two QTL were identified ($P < 0.001$) with no evidence of epistasis.

The plots in Figure 3 and Figure 5, and Tables 2–4 show that many of the associations between phenotype and genotype were repeatedly identified in the two independent RIL or NIL datasets, such as the area QTL on chromosome 1 in the 72.2–74.0 cM interval. In most but not all cases, Cvi DNA at the indentified positions had a positive effect on each of the traits relative to L*er* DNA, possibly indicative of genetic drift rather than natural selection (ORR 1998). This is shown by the allele effect plots in Figure 4. Only one case of epistasis seems to have been detected twice, appearing in RIL1 and RIL2. The specifics of this apparently repeatable instance of epistasis are shown in Figure 6. Figure 6A shows a locus at position 4.0 on chromosome 3 to have an effect on the area trait that depended on the genotype at position 19.6 on chromosome 1. Figure 6E shows what appears to be the same locus, position 8.3 on

chromosome 3, affecting the major axis trait to an extent that depends on the genotype at position 21 on chromosome 1. Other significant instances of epistasis are characterized in Figure 6B, C, D, F, and G.

Discussion

Fertilization of the egg within the Arabidopsis ovule produces the zygote, and a second fusion event involving a different sperm and the polar nuclei triggers differentiation of endosperm tissue, which in turn provides nutrients for the developing embryo. At maturity, the endosperm will be reduced to a single cell layer, the remainder being absorbed by the cotyledons and axis of the embryo to fund seedling growth and development before the initiation of photosynthesis. The axis contains shoot and root meristematic tissue that give rise to all parts of the mature plant. In Arabidopsis, all of this is ultimately packaged within an oval seed half a millimeter in length. Because the embryo closely fills the testa (Figure 1), the results presented here may be viewed as genetic maps of embryo phenotypes.

Benefits of the image processing technology

A feature of this study is the way in which automation made it feasible to include independent biological replicates of a RIL population and a NIL population without compromising sample size, resulting in more substantial evidence of multiple QTL than can be achieved by more typical phenotype analyses. The automation resulted by design from a good match between the capabilities of the custom analysis algorithm and the task of measuring the

morphologies of 10^2–10^3 individual, randomly scattered, sometimes touching seeds captured in an image ranging from 100 to 300 MB in size. Input files were readily processed in automated batch form, generating results with higher resolution than previous studies achieved. In fact, allele effects on the phenotype as little as 5% were resolved.

In addition to seed area, the algorithm measured seed length and width (major and minor axis, respectively). The multiple-QTL modeling of the measurements identified one locus unique to each of the two axis traits. A QTL on chromosome 2, position 41 was pertinent only to variation in the major axis, while one on chromosome 3, position 34 affected only the minor axis (Figure 5, Table 4). The major axis QTL on chromosome 2 was not detected by Alonso-Blanco *et al*. (1999) in their manual analysis of seed length, so it may be a dividend of the computational measurement method, or by Herridge *et al*. (2011) who used image analysis but with an algorithm that measured only seed area. Presumably, the major-axis effect on chromosome 2 was not large enough to affect area significantly or was compensated by an undetected reduction in minor axis. These comparisons with other studies and the internal comparisons made in Figure 5 indicate that area measurements capture most of the variation in length and width so that most seed size/shape QTL would be detected by measuring area only. However, by also measuring the two axes automatically, area effects could in most cases be ascribed to effects on width or length, or both (Fig. 5).

Differences in the two RIL populations

Although the two RIL populations were genetically identical, seeds in RIL2 were slightly larger than those in RIL1 (Figure 2). The RIL1 data set was obtained directly from seeds

provided by another laboratory at the University of Wisconsin and reared as described in Vaughn and Masson (2011). The seeds for the RIL2 data set were obtained as described in the Materials and Methods. Differences in maternal environmental parameters, such as light quantity, or watering schedules, as well as seed storage conditions before initiation of the experiments, may be responsible for the differences between RIL1 and RIL2 (ELWELL *et al.* 2011; MOUSSEAU and FOX 1998).

Relation to previous seed size studies

Chromosome 1 position 0 and chromosome 1 position 19 were identified by Alonso-Blanco *et al*. (1999) as affecting seed length, and both were also found using the area phenotype in the two RIL and the NIL generations in this study. Additionally, these loci were strongly supported using major and minor axis length as phenotypes in the three populations. An interval spanning these loci was also identified as an area QTL by Herridge *et al*. (2011). The QTL on the upper arm of chromosome 1 may be the *MEDEA* locus. MEA, a FIS-class protein subunit of the *Polycomb*-group complex, is a chromatin modifying enzyme that acts in the *FIS* group of genes that mediate seed development by repressing expression of target genes (CHAUDHURY *et al.* 2001; KÖHLER and GROSSNIKLAUS 2002; KÖHLER *et al.* 2003). Before fertilization, gene expression is only from the maternal allele, although the paternal allele functions after fertilization (GROSSNIKLAUS *et al.* 1998; KINOSHITA *et al.* 1999; LUO *et al.* 2000; YADEGARI *et al.* 2000). *FIS* genes are thought to be negative regulators of endosperm growth and development. *fis* mutants undergo seed development without fertilization, and the endosperm does not cellularize but enlarges during the later stages of seed growth (CHAUDHURY *et al.* 1997; KIYOSUE *et*

al. 1999). The locus at the top of chromosome 3 was identified in the present study as well as by Alonso-Blanco *et al.* (1999), but no candidate genes suggest themselves.

Both length loci found by Alonso-Blanco *et al.* (1999) on chromosome 4 were also found to affect major axis in this study. One of these loci has previously been predicted to be *APETALA2* (*AP2*) (JOFUKU *et al.* 2005). However, like Herridge *et al.* (2011), our methods resulted in a smaller LOD support interval that excludes *APETALA2* but could colocalize with *SHORT HYPOCOTYLS UNDER BLUE1* (*SHB1*). Lack of AP2 activity has been shown to impact seed development. Mutants in this gene have irregularly shaped seeds with increased amounts of seed proteins and fatty acids (JOFUKU *et al.* 1994; LEON-KLOOSTERZIEL *et al.* 1994; WESTERN *et al.* 2001). SHB1 has been shown to associate with the promoter regions of *MINI3* and *IKU2*, whose mutants show a reduced seed size (LUO *et al.* 2005; WANG *et al.* 2010; ZHOU *et al.* 2009), possibly through interactions with other proteins that bind to these regions. All three of these genes have been implicated in controlling the timing of endosperm growth and cellularization (ZHOU *et al.* 2009).

Of the two QTL on chromosome 5 found in the Alonso-Blanco study, both were also found in at least one of the populations analyzed here, with stronger evidence for the locus near the bottom of the chromosome. Three genes with known roles in seed size development colocalize with this distal QTL: *ARF2*, *TITAN3*, and *AGL62*. *ARF2* encodes a transcription factor that binds auxin-responsive elements in auxin-regulated genes' promoter regions (SCHRUFF *et al.* 2006; ULMASOV *et al.* 1999). Its role in cell proliferation was revealed as the integument of a mutant plant contains supernumerary cells without prior fertilization of the ovules (SCHRUFF *et*

al. 2006). *TITAN3* has been shown to be involved with the proliferation of endosperm nuclei early in seed development, although its mutants have mostly normal embryo development and undergo cellularization of the endosperm at the proper time (LIU and MEINKE 1998). *AGL62* helps to generate a mobile signal to initiate seed coat development, probably through interaction with type I MADS-box proteins such as PHERES1 (DE FOLTER *et al.* 2005; ROSZAK and KOHLER 2011).

The minor axis QTL on chromosome 3 appears to match the seed area QTL identified by Herridge *et al.* (2011) in a Bur x Col RIL population but, appropriately, it was not found by Alonso-Blanco *et al.* (1999) in their study of length. Figure 4 shows that substitution of Cvi alleles at this minor-axis locus had a positive effect on the phenotype in the RIL populations, but a negative effect in the NIL population. This disagreement could be due to two loci with opposite effects acting within the 1.5 LOD support interval found in the RIL population, whereas only one of these loci was segregating in the NIL population. Alternatively, this locus could act epistatically with another locus, but again lack of recombination in NIL could prevent the same results from being witnessed in this data set. Nonetheless, these RIL data indicate strong support for this locus acting to control the width of the seed, without an impact on the length.

QTL that affect the major axis may identify loci that contribute to the length of the embryo and some of those may play a general role in regulating plant size. The QTL on chromosome 2 contributing to variation in seed area and length but not width could, based on its position, be the *ERECTA* locus that is nonfunctional in the L*er* parent of the population used here (KOORNNEEF *et al.* 2004). The *ERECTA* locus, which is responsible for the smaller stature of

the adult in the Landsberg *erecta* accession was also found in a QTL analysis of developmental traits using the Col x L*er* RIL population (KEARSEY *et al.* 2003) and various correlated growth traits in a Kas x L*er* population (PRINZENBERG *et al.* 2010). Although the Cvi accession has relatively large seeds, its fresh weight and leaf length and width are consistently smaller than those of other accessions (STOKES *et al.* 2007).

Relation to previous studies of plant biomass

Several studies have found strong correlations between metabolic activity and biomass (CROSS *et al.* 2006; LISEC *et al.* 2008; MEYER *et al.* 2007; STEINFATH *et al.* 2010) so QTL that are involved in the variation in seed size could be due to metabolic activity within the seed influencing the size of the axis or the cotyledons. Research by Calenge *et al.* (2006) on the QTL associated with carbohydrate metabolism in different nitrogen environments found approximately 14 distinct loci that affect sugar concentration in the Bay-0 x Shahdara RIL population. One locus identified in an analysis of starch content in a nitrogen-rich environment, ST10.2, colocalizes with one of the QTL identified here on the distal end of chromosome 1. Another locus associated with starch and fructose content maps to the same region as a QTL identified here on the proximal end of chromosome 3.

Relation to previous studies of seed content QTL

Seed oils are an important part of human and animal nutrition, and also have uses in industrial applications. In many species, they are the fuel for the germinating seedling that enable establishment. Triacylglycerols are the main storage component for seed oils in many

species, and Arabidopsis provides an excellent model system for the commercially relevant *Brassica* oilseed crops due to its similar seed physiology and subsequent development. Arabidopsis, whose oil content is approximately 40% of seed dry weight, has successfully been used to probe natural variation of fatty acid content in seeds (LEMIEUX *et al.* 1990; MILLAR and KUNST 1999; O'NEILL *et al.* 2003). The sheer bulk of oil within the Arabidopsis seed makes it likely that some loci involved in variation in seed size may also be involved in controlling fatty acid content, although one study found no strong link between seed oil content and seed mass (HOBBS *et al.* 2004).

Indeed, two of the loci identified in our study colocalize with QTL involved in fatty acid content. A locus near the end of chromosome 2 is in the same region as a locus that accounts for 17% of the variation in seed oil content in the L*er* x Cvi RIL population (HOBBS *et al.* 2004), and could be due to the action of *FAD3*, which encodes a fatty acid desaturase present in the endoplasmic reticulum. Overexpression of FAD3 has been shown to decrease linoleic acid and increase linolenic acid content (SHAH *et al.* 1997). Mutations at this locus are semi-dominant, supporting the idea that this gene could contribute to variation in seed oil phenotypes. Further support for this locus was found in research utilizing numerous RIL populations from diverse environments (O'NEILL *et al.* 2012). However, this locus is also near that of *ERECTA*, and since our analyses only identified this region when using area and major axis, but not minor, as phenotypes, we believe it is probably *ERECTA* that is instead the underlying gene behind this QTL.

Another locus that has been identified as being involved in the control of seed oil content is that of *FAE1*, a 3-ketoacyl-Coa synthase that works to synthesis very-long-chain fatty acids (VLCFAs) in the endoplasmic reticulum of cells within the developing embryo. The proportions of VLCFAs in seed oil have been found to be quantitatively affected by transcription level of *FAE1* (MILLAR and KUNST 1997), and this locus has been identified in QTL analyses of seed oil quantity (O'NEILL *et al.* 2003; O'NEILL *et al.* 2012). *FAE1* is localized to the distal end of chromosome 4, where our study has consistently identified a region that contributes to variation in seed area, length, and width. If relationships between seed oil and some morphological feature detectable by image analysis can be established, the approach to phenotyping used here could be used to select genotypes with desirable chemical compositions.

Acknowledgements

We thank Patrick Masson, Department of Genetics, University of Wisconsin-Madison, for providing the initial set of L*er* X Cvi RIL seeds, and Karl Broman, Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, for much advice and critical reading of the manuscript. We thank Kelsey Rudd for assistance in harvesting seed. This work was supported by funding from the NSF Plant Genome Research Program (IOS-1031416).

Preface to Chapter 3

In Chapter 2, a high-throughput method of measuring seed size useful for studies of natural variation or mutant analysis was presented. In Chapter 3, I describe another phenotyping platform developed, in this case, for measurement of root gravitropism. The time course nature of this phenotype added additional layers of complexity not seen in our study of seed size. For one, the high temporal resolution used to capture the gravitropic response resulted in 241 individual time points, and thus phenotypes, for QTL analysis. This extensive set of data required enormous amounts of computing resources, especially for calculating significance thresholds. Partnering with HTCondor enabled these calculations and model selection procedures to occur over distributed computing networks, significantly decreasing the amount of time needed to obtain trait mapping results. Chapter 3 illustrates the results of these analyses, highlighting the time dependent nature of the identified loci. Surprisingly, the significant loci exhibited remarkably different time profiles, with some affecting variation in tip angle early, and some late, in the response. Determination of when a particular gene is acting during a response could provide clues as to how it functions within the plant. The rich dataset of root images freely available online can be further studied and acts as a model for how phenotype data should be shared with the research community.

Chapter 3:

High-throughput computer vision introduces the time axis to a quantitative trait map of a plant growth response

Abstract

Automated image acquisition, a custom analysis algorithm, and a distributed computing resource were used to add time as a third dimension to a quantitative trait locus (QTL) map for plant root gravitropism, a model growth response to an environmental cue. Digital images of *Arabidopsis thaliana* seedling roots from two independently reared sets of 162 recombinant inbred lines (RILs) and one set of 92 near isogenic lines (NILs) derived from a Cape Verde Islands (Cvi) x Landsberg *erecta* (L*er*) cross were collected automatically every two minutes for eight hours following induction of gravitropism by 90 degree reorientation of the sample. High Throughput Computing (HTC) was used to measure root tip angle in each of the 1.1 million images acquired and to perform statistical regression of tip angle against the genotype at each of the 234 (RIL) or 102 (NIL) DNA markers independently at each time point using a standard stepwise procedure. Time-dependent QTL were detected on chromosomes 1, 3, and 4 by this mapping method and by an approach developed to treat the phenotype time course as a function-valued trait. The QTL on chromosome 4 was earliest, appearing at 0.5 h and remaining significant for 5 h, while the QTL on chromosome 1 appeared at 3 h and thereafter remained significant. The Cvi allele generally had a negative effect of 2.6–4.0%. Heritability due to the QTL approached 25%. This study shows how computer vision and statistical genetic analysis by HTC can characterize the genetic architecture of developmental timing.

Introduction

Methodologies for characterizing genomes and genotypes are more advanced in terms of their degree of automation and throughput than their phenotype counterparts. This imbalance hinders progress in mapping genotype to phenotype in model systems such as the *Arabidopsis thaliana* plant, the subject of the present study. Computer vision methodologies can help address this imbalance by enabling objective, accurate, and potentially automated measurements of size and shape captured in digital images (SPALDING 2009). The Arabidopsis seedling root is well suited for this approach because it grows and responds well to stimuli in experimental scenarios that permit acquisition of high-quality images from which shape and size descriptors such as midlines, local curvature, and angles can be computationally extracted (CHAVARRIA-KRAUSER *et al.* 2008; FRENCH *et al.* 2009; MILLER *et al.* 2007). For example, Durham Brooks *et al.* (2010) used automated analysis of high-resolution images to measure the influence of seed size, seedling age, and media composition on the time course of root gravitropism, a dynamic growth response to a change in orientation with respect to the gravity vector (BLANCAFLOR and MASSON 2003; MORITA 2010). Miller *et al.* (2010) used the same approach plus extended computational analysis to quantify a transient gravitropism phenotype and its precise time of onset in young roots of *glr3.3* glutamate receptor mutants. Other researchers have taken advantage of automated image acquisition and analysis to measure shoot growth features over time (TISNÉ *et al.* 2013; ZHANG *et al.* 2012), or to acquire rotational series of branching root systems growing in a transparent gel to capture a third spatial dimension instead of time (CLARK *et al.* 2011; IYER-PASCUZZI *et al.* 2010).

The degree of automation made possible by computational image analysis increases the feasibility of systematically acquiring high-resolution phenotype data from genetically-structured populations for the purpose of large-scale statistical genetic studies. The benefits of accuracy and objectivity attending this approach increase the quality of the resulting phenotype data. Accordingly, automated or semi-automated image analysis is beginning to impact the mapping of quantitative trait loci in plants (HERRIDGE *et al.* 2011; MOORE *et al.* 2013; TISNÉ *et al.* 2013; TOPP *et al.* 2013). The project presented here combines an ability to capture the time course of gravitropism in high resolution with the throughput needed to make mapping populations feasible subjects. Gravitropism was selected for study because it is a model that embodies perception and transduction of an external cue, hormonal control of growth, and adaptation to a new environmental state (BOONSIRICHAI *et al.* 2002). These are fundamental physiological and developmental processes, now made accessible to statistical genetic analysis by novel application of automated computer vision and computation methodologies.

Materials and Methods

Germplasm

The Arabidopsis seed stocks used here were described in a recent machine-vision study of seed size and shape QTL (MOORE *et al.* 2013). As in the previous report, RIL1 refers to a set of 162 F_{10} RILs derived from a cross between the Landsberg *erecta* (L*er*) and Cape Verde Islands (Cvi) accessions of *Arabidopsis thaliana* (ALONSO-BLANCO *et al.* 1998) donated to us by Professor Patrick Masson, University of Wisconsin, and to the phenotype data set obtained with those

seeds. RIL2 refers to progeny of RIL1 produced in a random-designed and carefully controlled seed-bulking exercise in order to duplicate the population in a different maternal environment as described in Moore *et al.* (2013). NIL refers to 92 near isogenic lines created by introgression of various short regions of Cvi DNA into the L*er* background (KEURENTJES *et al.* 2007) and raised in the same manner as the RIL2 seeds, again as described in Moore *et al.* (2013). The plant populations or the phenotype data sets derived from them are referred to as RIL1, RIL2, or NIL.

Plant culture

In a row on a Petri plate containing 1 mM KCl, 1 mM $CaCl_2$, 5 mM 4-morpholineethanesulfonic acid, pH 5.7 with BIS-TRIS propane and gelled with 1% agar, three seeds were sown ~0.5 cm apart. Plates with seeds were placed at 4° for 2–4 d of stratification before being cultured vertically in a 22° growth chamber under constant white light for 3 d, at which point the primary root was 2–8 mm in length.

Root imaging

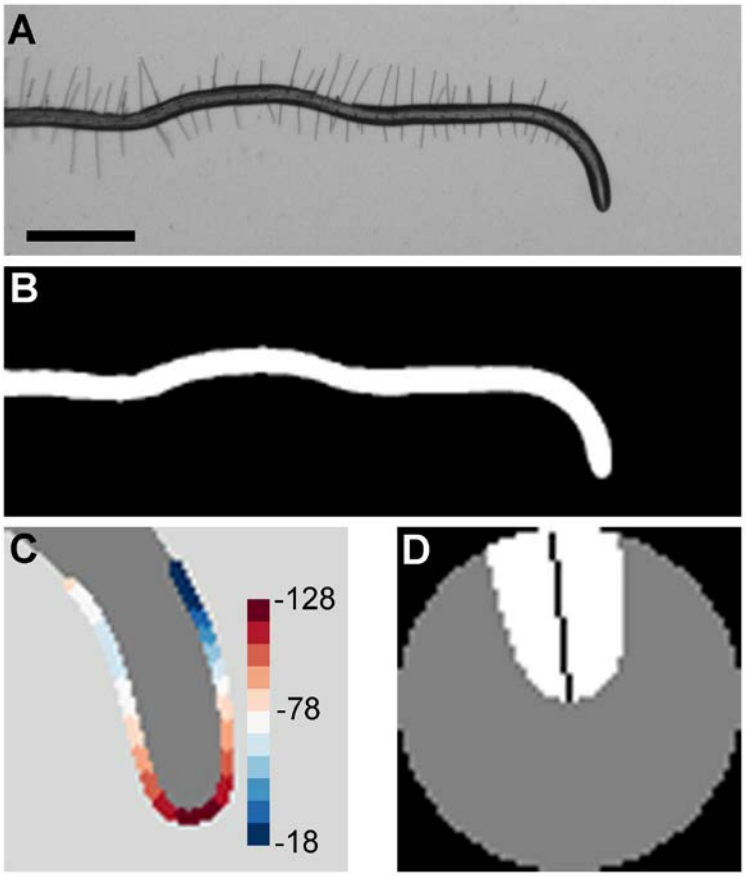
An imaging system consisting of 11 charge-coupled device cameras (Marlin F146B; Allied Vision Technologies) each equipped with a macro-zoom lens (model NT59-157; Edmund Optics) was used to capture digital images of roots on as many Petri plates, each backlit with 880 nm radiation in a room maintained at ~22°, essentially as previously described (DURHAM BROOKS *et al.* 2010; MILLER *et al.* 2010). Instructions on how to create a similar apparatus are posted at http://phytomorph.wisc.edu/hardware/fixed-cameras.php. To initiate the experiment, each Petri dish was held in a custom plate holder in front of a camera and rotated by 90°. Framing, focusing, and starting image acquisition, in that order, were performed within 20 s of sample

rotation. Images were acquired automatically every 2 min for 8 h at a resolution of 100 pixels/mm, resulting in a 241-frame movie of the root gravitropic response. The resulting images showed the dark roots of the three seedlings on a plate, against a light background. A video of an example experiment is provided as Supporting Information, File S1. Because each response recording lasted 8 h, two successive trials per camera were acquired within a given day to increase the throughput of phenotype acquisition. To ensure that the roots being compared were at similar growth stages, only those with initial lengths between 3 and 8 mm were used for QTL analysis. RIL1, RIL2, and NIL were phenotyped consecutively. Within each population, genotypes were selected for measurement according to a randomized order.

Tip angle measurement

The camera firmware that controlled acquisition also automatically saved the images to a local disk array. A file synchronization protocol nightly copied the images to disks housed in the Department of Computer Sciences from whence they were automatically read and processed for tip angle measurements on the University of Wisconsin Center for High Throughput Computing (CHTC) grid, which is managed and scheduled by the HTCondor software tool (THAIN *et al.* 2005). The custom algorithm that measured tip angle at each time point in a time series of frames was implemented in Matlab®. The algorithm first binarized the grayscale images to isolate the roots from the background (Figure 1A–B). Erosion and pruning produced a continuous, branchless skeleton of the root. The set of pixels forming the boundary of the root apex was located within an 81 x 81 pixel patch of the binarized root image centered at the terminus of the skeleton. The local curvature at each position of the boundary was

Figure 1: Root tip angle determined by image analysis. (A) Grayscale image of a representative root undergoing a growth response to a change in the gravity vector, or gravitropism. Scale bar = 1mm. (B) Binarized image of the responding root. (C) Curvature values are calculated at each boundary point along the root apex. The point of highest curvature is taken to be the root tip. Color legend indicates curvature values in mm^{-1}. (D) A patch centered at the root tip is subjected to principal components analysis. The first eigenvector (black line) determines the tip angle relative to the horizon.



determined by continuous wavelet transformation of a parameterized trace of the boundary pixels. The point of maximum curvature was labeled the root tip (Figure 1C). The largest circle fit within a 60 x 60 pixel patch centered at the root tip was used to mask (set to zero) a portion of the root apex large enough that the tip angle could be equated with the first principal component determined by principal components analysis (Figure 1D). The angle of each root tip in each frame was thus determined, stored in comma-separated value files, and automatically copied to the local server one day after the experiment was performed. The large majority of trials were successfully processed automatically, resulting in 8–20 independent time course measurements for each genotype. The most common problems encountered were blurry images caused by excessive condensation forming on the lid of the Petri plate, and two roots colliding during the experiment. In all, 2123, 2325, and 2536 roots were successfully imaged and analyzed in the RIL1, RIL2, and NIL populations, respectively.

Stepwise QTL analysis using HTCondor

To add a high-resolution time axis to a QTL map, stepwise QTL analyses (MANICHAIKUL *et al.* 2009) were performed at each of the 241 time points using tip angle as the phenotype. The approach uses a penalized LOD score, with the LOD score for a multiple-QTL model (the $\log_{10}$ likelihood ratio comparing the given model to the null model with no QTL) penalized by the complexity of the model with separate penalties for main effects and pairwise interactions derived by a permutation test with a nominal significance level of 5% in the context of a two-dimensional, two-QTL genome scan. The approach uses a penalty on the main effects, T_m, and

two different penalties on interactions, T_i^H (heavy) and T_i^L (light), and considers the structure of a QTL model to assign these penalties; see Manichaikul *et al.* (2009).

Performing 241 separate multiple-QTL analyses presented computational challenges especially because, to establish the penalties for each time point, 25,000 permutations of the phenotype against the genotype were performed in the context of a two-dimensional, two-QTL genome scan, using Haley-Knott regression (HALEY and KNOTT 1992). The calculations were performed with R/qtl (BROMAN *et al.* 2003), an add-on package for the R statistical software (R CORE TEAM 2013). A directed acyclic graph (DAG) created in DAGman (COUVARES *et al.* 2007) structured the permutation tests into 5 per job and the HTCondor software tool was used to execute the 1.2 million separate compute jobs (5000 jobs per time point x 241 time points) in parallel on the CHTC or the Open Science Grid (OSG) distributed computing resources. After the results of each permutation test were automatically collated, the DAG triggered automatic model selection and QTL processing, again performed using R/qtl (BROMAN *et al.* 2003). Model fitting was performed in the manner described in Moore *et al.* (2013).

Heat maps of profile logarithm of odds (LOD) scores were used to visualize the evidence for QTL in the context of a multiple-QTL model. This is analogous to a display technique used in Zeng *et al.* (2000): Each QTL was considered separately, and its position was allowed to vary, keeping all other QTL fixed at their estimated positions. For each possible position of the QTL under test, the LOD score comparing that multiple-QTL model to the model with the given QTL omitted was calculated.

Source code, raw and processed data

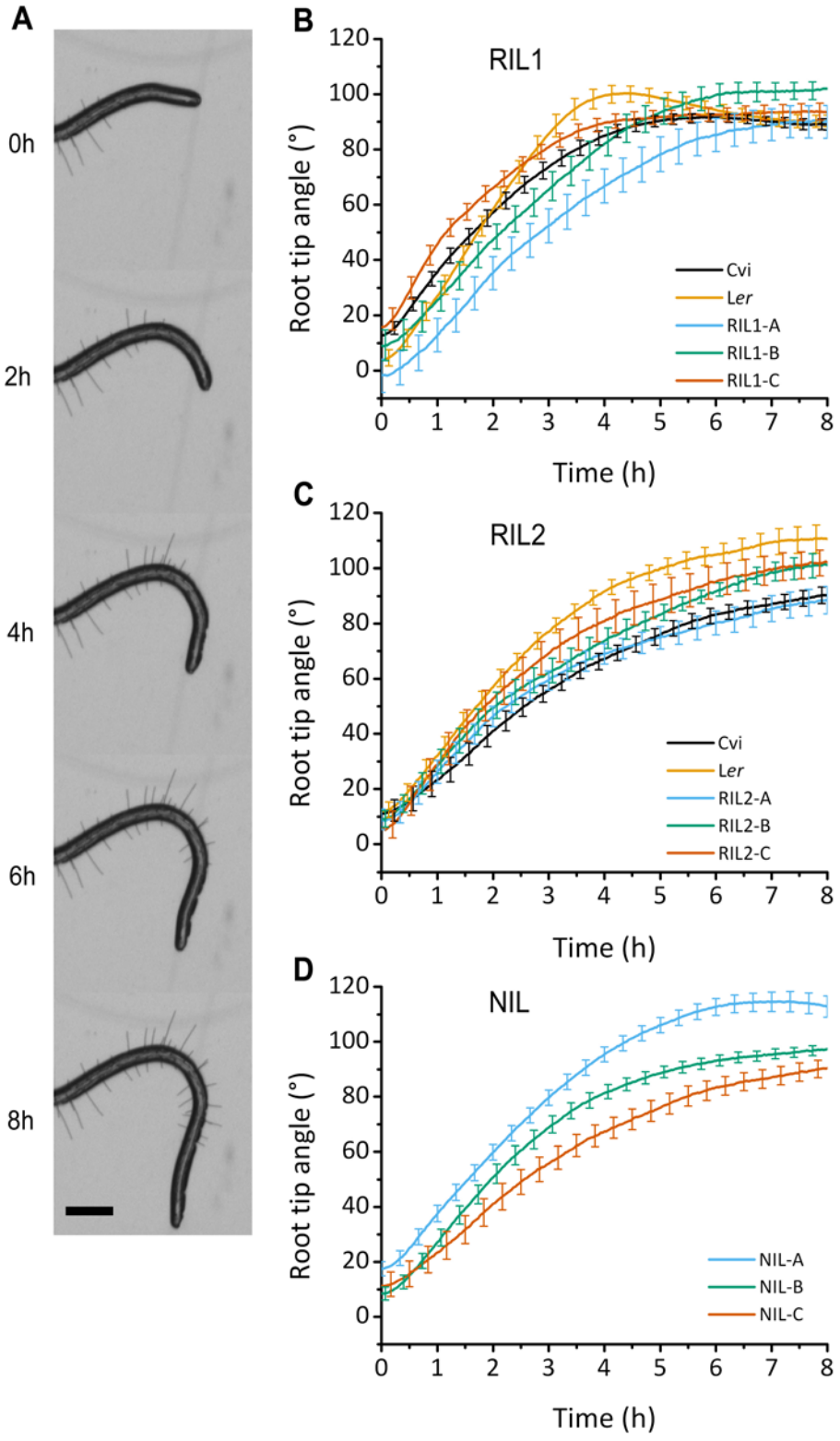
All of the raw images, custom image analysis computer code, quantified phenotype data, R/qtl input files, and reduced QTL results are available for download at http://phytomorph.wisc.edu/download/. These resources enable repetition of the analyses presented here but, probably more importantly, they enable computer vision scientists to extract additional traits from the image sets and statisticians to develop new methods for mapping phenotypes quantified with high time resolution.

Results

The gravitropism dataset

The biological material studied here was a set of 162 recombinant inbred lines of *Arabidopsis thaliana* derived from a cross between the Cape Verde Islands (Cvi) and Landsberg *erecta* (L*er*) ecotypes (ALONSO-BLANCO *et al.* 1998) and a set of 92 near isogenic lines containing introgressions of Cvi DNA in the L*er* background. The seed stocks used were the same previously studied by image analysis to determine the size distributions and genetic architectures of length, width, and area traits (MOORE *et al.* 2013). The phenotype data are time series of root tip angles extracted from sequences of grayscale digital images automatically acquired at 2 min intervals during gravitropism and batch-processed on a HTCondor-managed high-throughput computing grid (THAIN *et al.* 2005). Figure 2A shows five images of a representative root undergoing gravitropism for 8 h. Each of the 6,984 separate

Figure 2: Quantifying root tip angle during gravitropism generates a time-course phenotype. (A) A representative series of images of a root responding to gravity at 2 h intervals. At the beginning of the experiment, seedlings were rotated 90° such that the root tip was approximately horizontal to the gravity vector. By 8 h after rotation, the root had grown to reorient its tip parallel to the direction of gravity. Scale bar = 0.5 mm. (B–D) Tip angle development during gravitropism in the RIL1, RIL2, and NIL datasets. The response of the Cvi parental line is shown by a black line, L*er* by an orange line, and the other lines indicate the responses of randomly selected representatives of the population of inbred lines (labeled A, B, and C). Vertical bars indicate the standard error of the mean. For RIL1, n = 27, 28, 10, 10, and 18 for Cvi, L*er*, RIL1-A, RIL1-B, and RIL1-C. For RIL2, n= 14, 18, 18, 11, and 12 for Cvi, L*er*, RIL2-A, RIL2-B, and RIL2-C. For NIL, n = 32, 33, and 20 for NIL-A, NIL-B, and NIL-C.

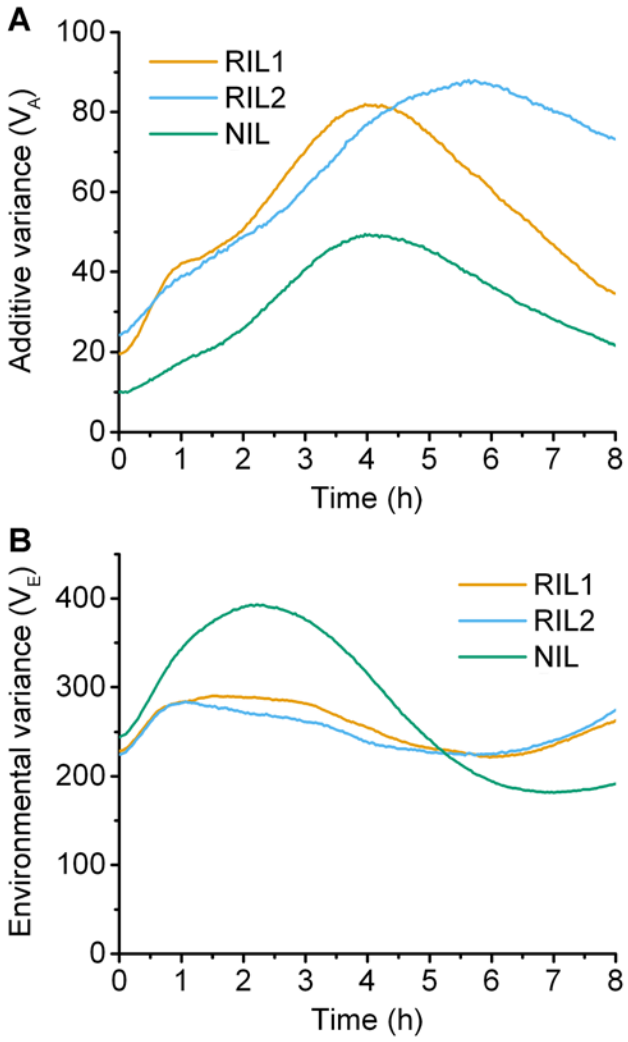

A
0h
2h
4h
6h
8h
B
RIL1
Root tip angle (°)
Time (h)
Cvi
Ler
RIL1-A
RIL1-B
RIL1-C
C
RIL2
Root tip angle (°)
Time (h)
Cvi
Ler
RIL2-A
RIL2-B
RIL2-C
D
NIL
Root tip angle (°)
Time (h)
NIL-A
NIL-B
NIL-C

trials in this study consists of 241 frames that form a time series. The average tip angle responses of the two parental ecotypes and three example members of the mapping population are shown in Figure 2B. Regardless of genotype, all roots reoriented by approximately 90° within 8 h though the time courses differed significantly. The set of phenotype data obtained for this set of recombinant inbred lines is referred to as RIL1. A separate rearing and harvest of the recombinant inbred lines including the parental accessions and a repeat of the phenotype measurements from the next generation of seedlings produced the RIL2 dataset. In general, RIL2 responses were slower than RIL1 responses despite the genetic identity of the plants (Figure 2C). The same measurements were performed on the near isogenic lines, the seeds of which were generated in the manner of the second RIL generation, to produce the NIL dataset (Figure 2D).

Analysis of variance

The variance attributed to additive genetic variation and environmental variance in RIL1, RIL2, and NIL are plotted in Figure 3. Additive genetic variance was initially low in all populations but increased as the gravitropic response progressed, peaking 4 to 6 h after gravistimulation (Figure 3A). Environmental contribution to total phenotypic variance exceeded the variance attributed to additive genetic variation, developing mostly during the first few hours of the response, especially in NIL (Figure 3B). These analyses of variance are based on the differences between the average responses of genetically distinct lines.

Figure 3: Genetic and environmental contributions to variance. (A) The proportion of the phenotypic variance attributed to the additive genetic variation. (B) The environmental contribution to the phenotypic variance. Orange, blue, and green lines indicate the RIL1, RIL2, and NIL populations, respectively.



Time-resolved QTL map

The mean tip angle at each of the 241 time points for each genotype was used as the phenotype in 241 separate and independent multiple-QTL modeling analyses. The result was a highly time-resolved QTL map for each of the datasets (Figure 4A−C). LOD score is shown on a color-intensity scale as a function of time (abscissa) and genome position (ordinate). In RIL1 (Figure 4A), a total of 8 loci, with a maximum of 6 loci at any given time, were detected during the 8 h response. Two loci in particular were present during a large portion of the response. The locus on chromosome 4 at 40.3 cM affected variation in the response from approximately 0.5 to 6 h past gravistimulation, and the other on chromosome 1 at 64 cM contributed from approximately 3 to 8 h. Some QTL were more limited in duration. For example, shortly after 3 h a QTL appeared on chromosome 3 at 44 cM and lost influence at 5 h. This particular QTL illustrates a problematic feature of treating the phenotype at one time point independent of the next, namely that a QTL can abruptly disappear or shift position due to data at adjacent time points being best fit by different QTL models.

RIL2 supported the selection of 3 loci as genetic contributors to variation in the gravitropism response (Figure 4B). Two loci contributed from 2 to 8 h after gravistimulation, with the third locus on chromosome 3 at 17 cM arising later to influence the response between 4 and 8 h after the start of the experiment. A locus on chromosome 4 at 40 cM was identified and displayed a similar time course in RIL1 and RIL2 (Figure 4B).

Figure 4: Time course of QTL development. Magnitude of LOD score is displayed as color intensity as a function of time. (A–C) Profile LOD scores for models selected by stepwise QTL analyses, considering each time point individually. For RIL1, (T_m, T_i^H, T_i^L) = (2.58, 3.41, 1.53), for RIL2, (T_m, T_i^H, T_i^L) = (2.56, 3.44, 1.63), and for NIL, (T_m, T_i^H, T_i^L) = (2.75, 2.40, 0.67). (D–F) Single-QTL analysis results from RIL1, RIL2, and NIL populations. Threshold for significance of single QTL is 1.91 for RIL1, 1.99 for RIL2, and 1.91 for NIL. (G–I) Profile LOD scores for the model selected by stepwise QTL analyses using the SLOD criterion. For RIL1, (T_m, T_i^H, T_i^L) = (1.91, 2.41, 1.10), for RIL2, (T_m, T_i^H, T_i^L) = (1.99, 2.62, 1.51), and for NIL, (T_m, T_i^H, T_i^L) = (1.91, 2.40, 1.66). The position of the loci on the y-axis is shown as the cumulative position of the five Arabidopsis chromosomes. Horizontal lines indicate chromosome breaks. The x-axis shows time in hours since onset of gravistimulation. All coordinates whose LOD scores were not significant are shown in white. Darker blue colors indicate higher support for the loci.

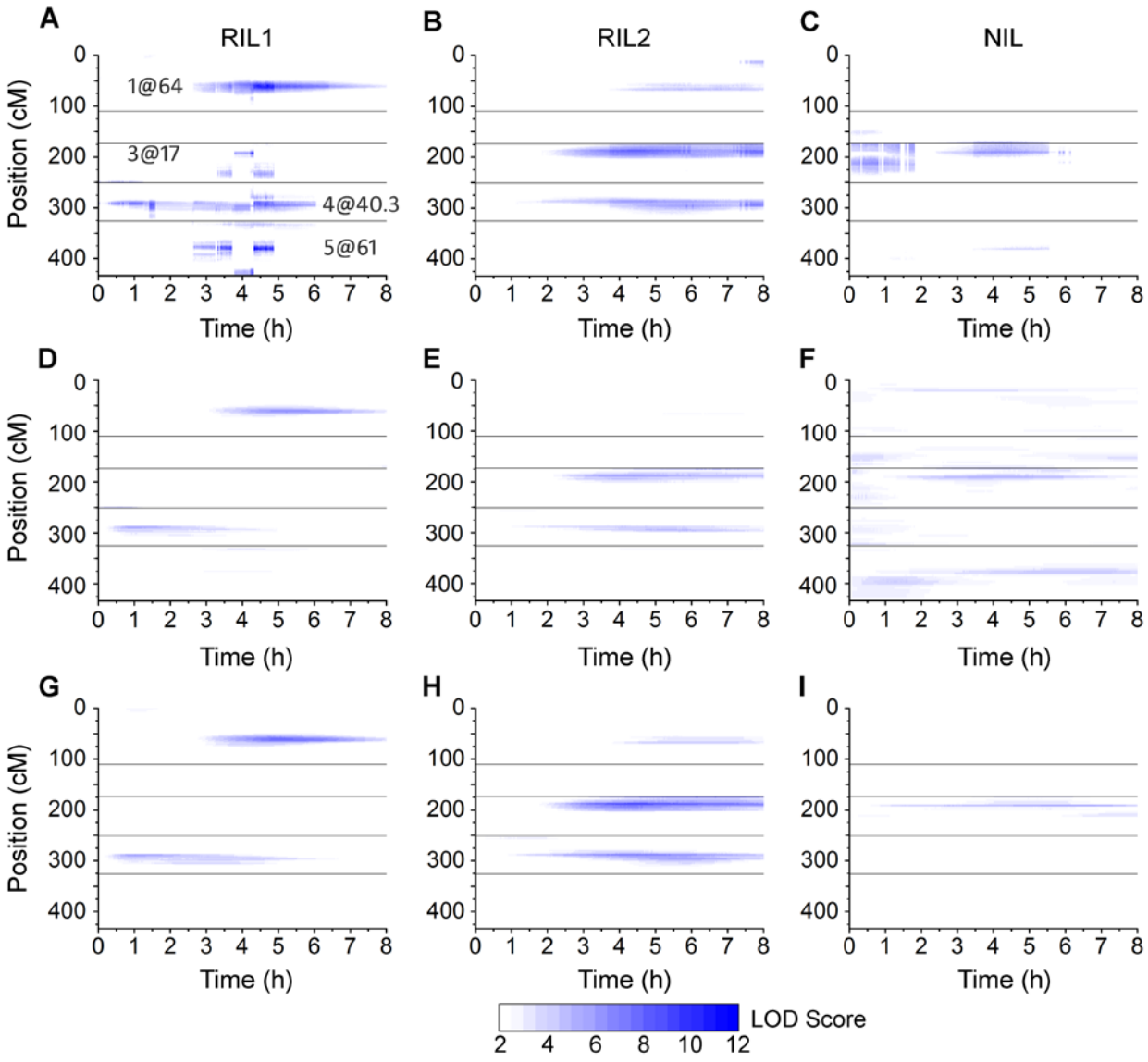

A
RIL1
B
RIL2
C
NIL
D
E
F
G
H
I
Position (cM)
Time (h)
1@64
3@17
4@40.3
5@61
LOD Score
2 4 6 8 10 12

Figure 4C shows that 3 loci were statistically significant in NIL during the early and middle parts of the response. The locus on chromosome 3 at 17 cM detected throughout the first 6 h of the response was also featured prominently in RIL2, and in the 'alternate' model centered around 4 h in RIL1. A second locus on chromosome 5 at 55 cM, influential from 3.5 to 5.5 h after gravistimulation, was also apparent over the same time period in RIL1. NIL was the only dataset to support a QTL on chromosome 2, located at 40 cM, which contributed to the variation in the response during the first hour.

A function-valued approach

The analysis described above adjusted for the search across QTL models but not for the search across time points. Another shortcoming is that the selected model occasionally abruptly changed between adjacent time points despite the phenotype following a smooth curve across time. A simple but effective approach to integrating information across time points was devised to accommodate the high time resolution in the phenotype data generated by the automated image analysis. The method used a model comparison criterion analogous to that of Manichaikul *et al.* (2009). It was initiated by performing a genome scan, or single-QTL analysis, at each time point to generate a 2D matrix of LOD scores (genomic positions x time points). Figure 4D–F shows the results of these single-QTL analyses in the three populations. For RIL1, strong support was obtained for loci on chromosomes 1 and 4 acting at different times (Figure 4D), while for RIL2, loci affecting this response were predicted on chromosomes 3 and 4 (Figure 4E). The locus on chromosome 4 was predicted to be at approximately 40.3 cM in both

RIL1 and RIL2. The NIL population showed some support for a locus on the proximal end of chromosome 4 and some support for loci on chromosomes 3 and 5 (Figure 4F).

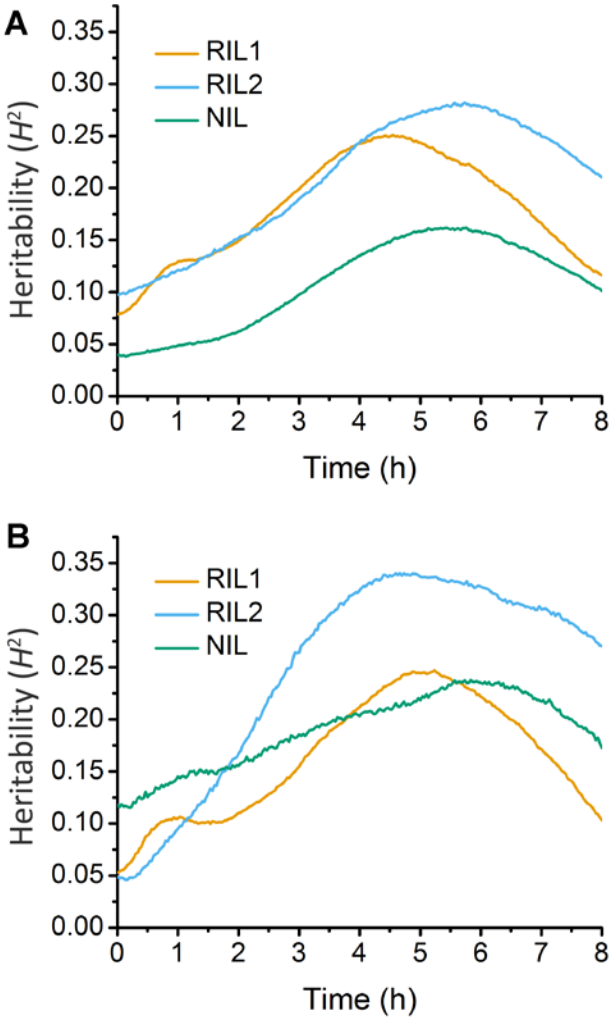
The next step in the method was to calculate the maximum LOD (MLOD) and the average LOD (SLOD) at each genomic position across time. Significance thresholds were obtained by a permutation test (CHURCHILL and DOERGE 1994), using the corresponding SLOD or MLOD statistic. SLOD exhibits higher power to detect QTL with effects across a large time interval, while MLOD exhibits higher power to detect QTL with large effects over a narrow interval. The MLOD or SLOD statistics were used to derive a penalized LOD criterion, which multiple-QTL analyses sought to maximize in a stepwise search, starting with the first QTL at the position with the highest MLOD or SLOD score. Additional QTL were added to the model if the significance threshold was met. After selecting the QTL model with the highest penalized LOD score, profile LOD scores were derived to evaluate the evidence and localization of each QTL, evaluating each time point in the 8 h response, individually: The position of each QTL was varied, one at a time, and at each location for a given QTL we derived a LOD score comparing the multiple-QTL model to the model with the given QTL omitted. Figure 4G–I present these profile LOD scores, based on the model identified with the stepwise QTL analysis using the SLOD statistic. In RIL1, a 2-QTL model was identified using the SLOD statistic, with loci on chromosomes 1 and 4 (Figure 4G), and a 2-QTL model with evidence of epistasis was predicted using the MLOD statistic, with the same QTL on chromosome 4 and an additional one on the distal end of chromosome 3 (data not shown but presented for download). For RIL2, the SLOD statistic predicted a 3-QTL model, with loci on chromosomes 1, 3, and 4 (Figure 4H), while the

MLOD statistic gave a 4 QTL model, with two loci on chromosome 1 and one locus on chromosomes 3 and 4 (data not shown but presented for download). In the NIL dataset, the SLOD statistic resulted in a 2-QTL model with both loci on chromosome 3, at 16.4 and 39.7 cM (Figure 4I). Analysis of this population using the MLOD statistic found both the two QTL on chromosome 3, as well as an additional locus on chromosome 2 at 81.3 cM (data not shown but presented for download).

Heritability

Analysis of variance was performed to calculate the heritability (H^2) due to the identified QTL across time in each population, using both the stepwise QTL analysis results at individual time points as well as the results achieved via stepwise QTL analysis using the SLOD statistic. For RIL1, H^2 ranged from a low of 10%, occurring at the initiation of the gravitropic stimulus, to a high of 27% around 6 h post-gravistimulation using trait data from each time point (Figure 5A). Analysis of RIL1 with the SLOD statistic showed a similar range in H^2, from 5% to a high of 33% (Figure 5B). The heritability calculated from the RIL2 population was somewhat lower than RIL1, ranging from 8–24% and 5–23% for analysis at each time point and via the SLOD statistic, respectively (Figure 5A–B). For the NIL population, the measured heritability was less than that of the RIL1 and RIL2 populations, with estimates of 3–15% from individual time point data (Figure 5A) and 12–22% from the SLOD statistic analysis (Figure 5B). The lower heritability of NIL can probably be attributed to the lesser amount of genetic variation present in these lines.

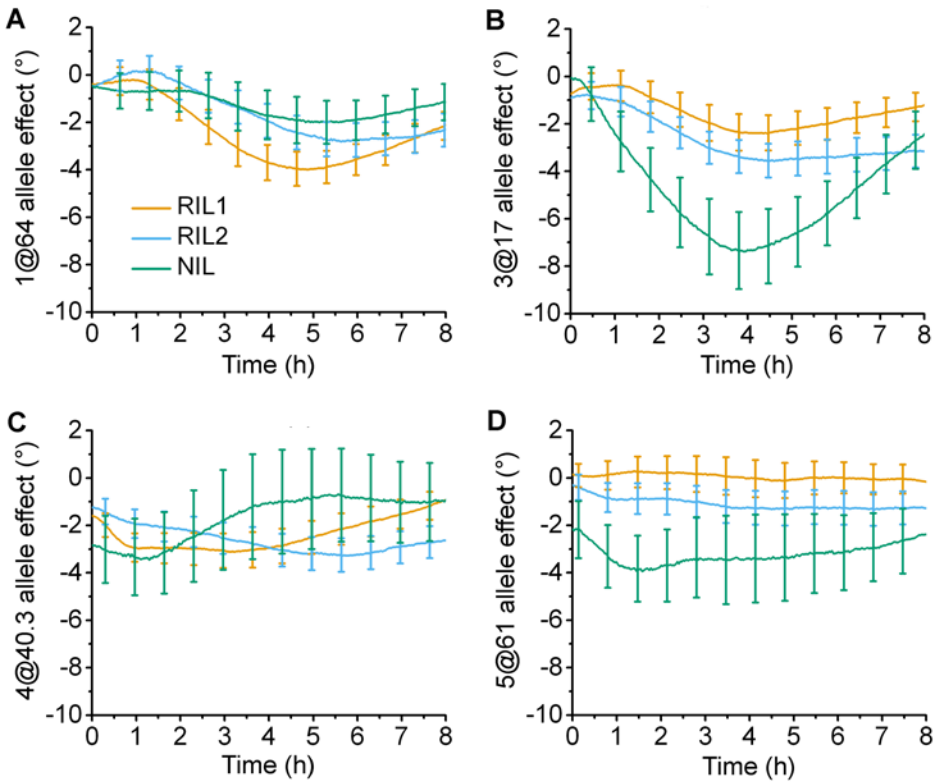
Figure 5: Heritability estimates from gravitropic responses of the L*er* x Cvi RIL and NIL populations. (A) Heritability calculated via results from stepwise QTL analysis at individual time points. (B) Heritability based on the chosen SLOD model from each population. Orange, blue, and green lines indicate the RIL1, RIL2, and NIL populations, respectively.



A four QTL model

Independent analysis of the three data sets identified some shared and some unique QTL. In an effort to consolidate the results in a single analysis, loci at chromosome 1 at 64 cM (1@64 cM), 3@17 cM, 4@40.3 cM, and 5@61 cM, identified as contributors by stepwise QTL analysis at each time point (Figure 4A–C), were used to create a four-QTL model that was evaluated in each of the three data sets. Figure 6 shows plots of the effect on the tip angle resulting from replacement of the L*er* allele with the Cvi allele at the indicated positions, i.e. the Cvi allele effect, as inferred from the four-QTL model fit separately to each of the data sets. Figure 6A shows that the Cvi allele at the 1@64 cM locus was estimated to have a negative time-dependent effect on the root tip angle in all three populations. The LOD scores for this position (Figure 4A–C) indicate strong support for the negative effect of this allele in RIL1 and RIL2, peaking around 5 h. Locus 3@17 cM was also estimated to have a negative effect on root tip angle, this time with NIL providing strong support, peaking about 4 h into the response (Figure 6B). The other two loci, 4@40.3 cM and 5@61 cM, also showed a negative Cvi allele effect but with inconsistent (Figure 6C) or little (Figure 6D) time dependence.

Figure 6: Time-dependent allele effects on root tip angle during gravitropism. The four QTL found by stepwise QTL analyses at each time point that were common to at least two populations were added to a model and the contribution of each locus was estimated. Positive values indicate that substitution of a Cvi allele at the indicated locus increases the tip angle trait, while a negative value corresponds to the L*er* allele increasing the trait value. The 95% confidence interval for the degree of the effect is displayed as error bars. (A) Chromosome 1 at 64 cM. (B) Chromosome 3 at 17 cM. (C) Chromosome 4 at 40.3 cM. (D) Chromosome 5 at 61 cM.



Discussion

Gravitropic signaling initiates in specific cells of the root cap (HASHIGUCHI *et al.* 2013), resulting in a redistribution of indole-3-acetic acid (IAA) flowing back shootward such that a growth-inhibiting level of this auxin accumulates on the lower side of the root (SPALDING 2013). Slowing of cell expansion on the lower side drives downward bending, which is first detectable within approximately 15 minutes (LEWIS *et al.* 2007; MILLER *et al.* 2007). The response time course in Arabidopsis seedling roots depends on factors such as seedling age, the size of the seed from which the seedling emerged, nutrient conditions (DURHAM BROOKS *et al.* 2010), and even on the previous generation's culture environment (ELWELL *et al.* 2011) but, generally speaking, the bending speed increases during the first two hours, peaks, then gradually slows as the tip angle approaches the new vertical. The present study captured the time course of this process across genetically structured populations of Arabidopsis seedlings so that the dynamics of its genetic architecture could be determined.

HTC advantages matched technical needs of temporal QTL study

Automated image acquisition and feature extraction were important contributors to the feasibility of this study but were not sufficient for success. Also necessary was the establishment of an automated HTC workflow. HTCondor software enabled batch processing of raw images and statistical genetic modeling of the quantified feature by matching jobs and available resources in a distributed computing environment. Obtaining the results presented in Figure 4A–C alone entailed launching 21,700 jobs that consumed approximately 5000 central processor unit hours. Such work would probably not have been feasible in the absence of such

a computing infrastructure. Data quality was also increased by adopting HTC because the degree of throughput and automation increased the feasibility of larger sample sizes.

Comparison of the two methods of QTL analysis

The first of the two methods of searching for significant QTL used in this study relied on automated stepwise QTL selection of a model based on tip angle data at a single time point, while the second method combined information from all time points to choose a multiple-QTL model. The first method is conceptually simpler, treating each time point independently and separately as a new trait to map. However, especially in the case of RIL1, the independent treatment of time points produced occasional discontinuities along the time axis as the selected QTL model jumped between alternatives. The alternate QTL positions, for example around 4 h on chromosomes 3 and 5 (Figure 4A), could reflect genes that contribute to the response but without sufficient strength to force their selection in each independent modeling exercise. In the second method, the function-valued approach, one model is chosen to represent the entire response, leading to improved power to detect QTL and better separation of linked QTL. The second approach is also more conservative as it seeks to control for the search across time points as well as the search across QTL models. When applied to the two RIL data sets, both methods detected the QTLs on chromosomes 1, 3, and 4 (Figure 4). Replacement of L*er* with Cvi alleles at these loci had a negative effect on the root tip angle (Figure 6) as did the Cvi allele at the weakly supported QTL on chromosome 5, which was only detected with the first method and only in RIL1 (Figure 4A).

Similarities and differences between RIL1 and RIL2

Independent rearings of the same set of recombinant inbred lines created the seed stocks used to produce the RIL1 and RIL2 data sets. Therefore, differences between the two data sets cannot be due to genetic differences. Instead, environmental differences during the production of the seed lots are the most probable causes of the differences between RIL1 and RIL2 root responses. Variation in seed size and shape within the two stocks were previously measured by image analysis and subjected to statistical genetic analysis. The seeds used to generate RIL2 were found to be significantly larger than those which generated RIL1. Such maternal environment effects on Arabidopsis seed size are known to affect next-generation growth of seedlings including root gravitropism (DURHAM BROOKS *et al.* 2010; ELWELL *et al.* 2011). Therefore, the QTL on chromosome 3 that is apparent in RIL2 (Figure 4H) but only weakly and briefly or not at all in RIL1 (Figure 4A,G) may be a locus that affects gravitropism only in certain seed-size contexts. However, none of the seed size or shape QTL previously identified (MOORE *et al.* 2013) using the same seed stocks had 1.5 LOD support intervals that overlapped with those of the root tip angle QTL mapped here.

Often, complex traits are affected by environmental factors in addition to the influence of the genetic architecture of the population (AHMADIYEH *et al.* 2003; SOLBERG *et al.* 2004). Further analysis of the data sets presented in this thesis could include the use of covariates to reduce the residual variation of the phenotypes and may increase the ability to locate QTL. For the seed size data (MOORE *et al.* 2013) detailing the QTL affecting variation in seed morphology of the material used in this study, using the identities of the mother plants that produced the

seeds as covariates may allow separation of loci only involved in the variation in seed size from those involved in QTL x environment interactions. For this gravitropism response data, the mother plant identities, the average seed size of each inbred line, the initial root length, or the initial tip angle of each root would also be potentially useful covariates. However, when using secondary phenotypes as covariates that are not always independent of genotype, caution should be taken in interpreting the results as the QTL identified might be representative of genetic factors involved in QTL x environment interactions during root gravitropism, rather than loci with an environment-independent effect (ZEEGERS *et al.* 2004).

Another way to uncover the underlying factors controlling variation in seed size and the root gravitropic response would be to perform joint mapping in the manner of Jiang and Zeng (1995). Here, the authors describe methods for formal testing of pleiotropic QTL effects, QTL x environment interactions, and the same trait from samples grown in different environments (see FALCONER 1952) which take into account the correlated structures of the data. These joint analyses may improve the power to find statistically significant QTL and increase the precision of the location and effect of these loci (JIANG and ZENG 1995).

Candidate genes

The 1.5-LOD support intervals calculated for each QTL permitted consideration of whether or not any corresponded to genes known to encode components of the gravitropism mechanism, such as PIN2 auxin transporters (CHEN *et al.* 1998; MÜLLER *et al.* 1998), determinants of starch content in the sedimenting statoliths (CASPAR and PICKARD 1989; KISS *et*

al. 1989), and plastid outer envelope proteins (STANGA *et al.* 2009), among others (BALDWIN *et al.* 2013; MORITA 2010).

The 1.5 LOD support interval of QTL 1@61 spans a region approximately from gene AT1G21945 to gene AT1G22420. One gene within this region is *AUXIN UP-REGULATED F-BOX PROTEIN2* (*AUF2*), which functions with the related AUF1 protein to regulate root growth responses to the hormones cytokinin and auxin (ZHENG *et al.* 2011). Mutating *AUF1* and *AUF2* impairs both of the major auxin transport streams in the root and makes root growth more sensitive to auxin transport inhibitors (ZHENG *et al.* 2011). Natural variation at this genetic locus could potentially affect the gravitropism time course.

The 1.5 LOD support interval for QTL 3@15 spans from AT3G23980 to AT3G24400. No gene within this region is known to function in gravitropism, hormonal control of root growth, or other processes that could support a hypothesis about how natural variation within it could affect the phenotype measured here. However, two genes within this interval, *MYB71* (AT3G24310) and *RIN4* (AT3G25070), show differential expression between the L*er* and Cvi accessions (VAUGHN and MASSON 2011).

The 1.5 LOD support interval for QTL4@40.3 extends from approximately AT4G15080 to AT4G15396 and includes *ABCG30*, a gene predicted to encode a membrane protein belonging to the ATP-binding cassette (ABC) superfamily of transporters (VERRIER *et al.* 2008). Members of the B subfamily of ABC transporters have been linked to IAA transport (NOH *et al.* 2001) and subsequently shown to affect gravitropism (LEWIS *et al.* 2007; NOH *et al.* 2003), but only relatively recently have members of the G subfamily been connected to auxin transport. Specifically, ABCG36 and ABCG37 have a subcellular localization and an apparent transport

activity that leads to efflux from roots of indole-3-butyric acid (RŮŽIČKA *et al.* 2010; STRADER and BARTEL 2009), a compound that has intrinsic auxinic activity and is believed to be a precursor of IAA (STRADER and BARTEL 2011). *ABCG30*, formerly known as *PLEIOTROPIC DRUG RESISTANCE 2* (*PDR2*), resides near *ABCG36* and *ABCG37* in the *ABCG* phylogenetic tree, and it is expressed specifically in roots (VAN DEN BRULE and SMART 2002). Expression in roots and a close sequence relationship with membrane proteins now believed to function in auxin homeostasis makes *ABCG30* an interesting candidate for the cause of phenotypic variation associated with QTL 4@40.3. Another gene within this interval, a UDP-glycosyltransferase superfamily protein (AT4G15260), is strongly downregulated in L*er* as compared to Cvi (VAUGHN and MASSON 2011). When the UDP-glycosyltransferase from *Pisum sativum* is expressed in Arabidopsis, gravity sensing is reduced in the root tissue (WOO *et al.* 2007).

The genes within the 1.5 LOD support interval surrounding QTL 5@61, covering loci AT5G16770 to AT5G17330, include *MERISTEM-DEFECTIVE* (*MDF*), which encodes a protein in the SART-1 family. MDF is expressed throughout the plant and mediates root, shoot, and flower development (CASSON *et al.* 2009). Messenger RNA levels of *PIN2* and *PIN4* are lower in *mdf* mutants, affecting the ability of the root to achieve a normal auxin maximum in the meristem, which may be how MDF functions to maintain meristematic activity (CASSON *et al.* 2009). The influence of MDF on meristem activity does not make it a strong candidate for the causative gene at this QTL because cell division is not expected to play an important part during an 8 h gravitropic response. However, the effect of MDF on PIN2 expression makes it a very interesting candidate of causation. PIN2 mediates the shootward flow of auxin that is an important component of the mechanism that creates differential growth between the upper

and lower sides of the root during gravitropism (SPALDING 2013). Natural variation in MDF may cause natural variation in PIN2 expression, and thereby affect the strength of the shootward auxin transport stream and the tropic responses that depend on it.

Because the gravitropic response is an important contributor to the phenomenon of root skewing, the growth pattern resulting from the deviation of a seedling root axis from the gravity vector when grown on a vertically maintained agar surface, one might expect that this study and the work of Vaughn and Masson (2011) would have found common significant loci. However, none of the loci from these two studies had overlapping LOD support intervals, probably due to the different phenotyping platforms that were used. In Vaughn and Masson (2011), seedlings were grown on media containing a higher agar concentration (1.5%), and initiation of the assay was begun with a 30° tilt of the previously vertical Petri dishes. Also, the degree of root skewing was measured after 2 days of growth. These differences in growth conditions and in the time course of the measurements, as well as the addition of the touch response in the phenotyping setup by Vaughn and Masson (2011), could explain why no loci were found in common between the two experiments.

Adding a highly-resolved time dimension to an analysis of quantitative trait loci posed some significant technical challenges met here by an emphasis on automated data acquisition and HTC for data analysis and modeling, but it also generates significant new insights and perspectives on the genetic architecture of the process that gives rise to a trait, rather than a trait itself. With image analysis becoming more common in studies of plant biology, in

scenarios ranging from confocal microscopes to whole plants (SPALDING and MILLER 2013), time may be an axis on many more QTL profiles in the future.

Acknowledgements

This work was supported by National Institutes of Health grant GM074244 to K.W.B. and by grant IOS-1031416 from the National Science Foundation Plant Genome Research Program to E.P.S. The authors Il-Youp Kwak and Karl W. Broman created the SLOD and MLOD methods of finding time-dependent QTL independent of the author of this thesis.

Appendix

This appendix describes additional characteristics of the populations studied in this thesis and information concerning how experimental parameters of the gravitropism experiment may influence the resulting growth response measurements, as well as discussion of additional analyses that can be performed using this data.

Due to the large number of genotypes analyzed in this study, human error could cause some seed tubes to be mislabeled, thus affecting the QTL analysis. To confirm the absence of labeling errors, average seed traits are plotted for RIL2 versus RIL1 in Figure 1. The strong correlations present for the area (Figure 1A), major axis (Figure 1B), and minor axis (Figure 1C) phenotypes suggest a small chance that the samples were incorrectly named. For example, the inbred lines in the RIL1 population that have large seed sizes also have large seed sizes in the RIL2 population. These graphs also illustrate that seeds from the RIL2 harvest are generally larger than those from RIL1.

Because seed area is a function of both length and width, as the measured length or width increases, we expect seed area to also increase. This phenomenon can be seen in Figure 2. For both the RIL1 and RIL2 populations, area is positively correlated with both major axis (Figure 1A) and minor axis (Figure 1B). Major axis and minor axis lengths are also positively correlated, but to a lesser degree (Figure 1C). The fact that the major and minor axis lengths are not completely dependent on each other perhaps allowed the identification of QTL that are only involved in the variation in one of these traits (see Chapter 2). As in Figure 1, these graphs also reveal the increased seed sizes of the RIL2 population compared to those of RIL1.

Figure 1: Comparison of seed size traits between the RIL1 and RIL2 populations. The average values of the (A) area, (B) major axis, and (C) minor axis of each genotype in the RIL1 and RIL2 populations are compared. The best fit line calculated via linear regression is plotted in red.

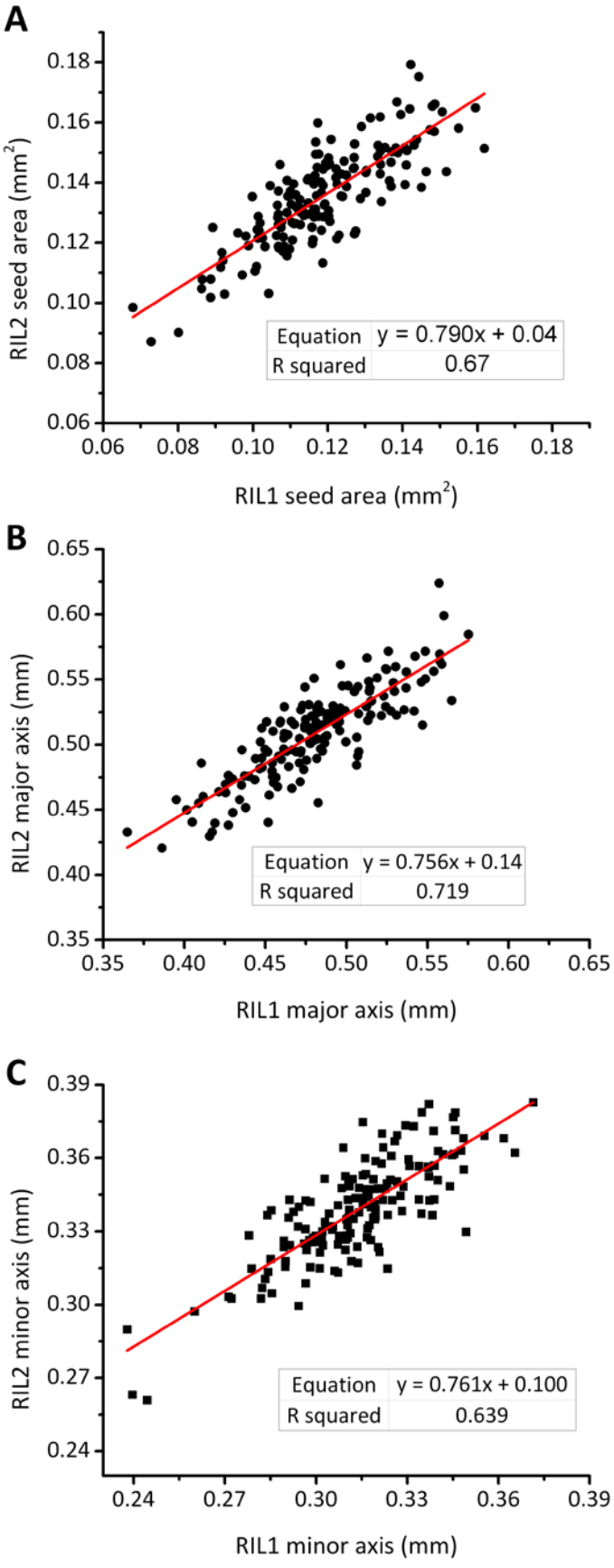
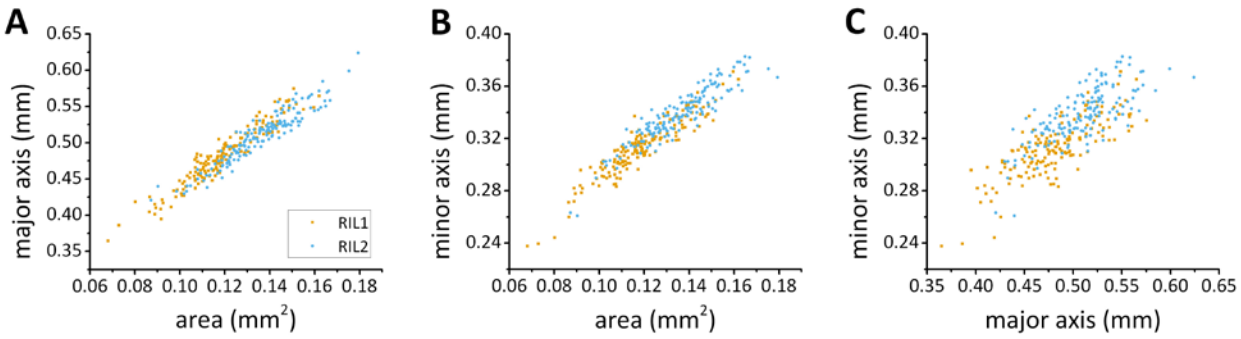

A
RIL2 seed area (mm²)
RIL1 seed area (mm²)
Equation y = 0.790x + 0.04
R squared 0.67
B
RIL2 major axis (mm)
RIL1 major axis (mm)
Equation y = 0.756x + 0.14
R squared 0.719
C
RIL2 minor axis (mm)
RIL1 minor axis (mm)
Equation y = 0.761x + 0.100
R squared 0.639

Figure 2: Correlations between seed size traits in the RIL1 and RIL2 populations. Average trait values of each genotype in RIL1 (orange) and RIL2 (blue) are plotted. (A) major axis vs. area, (B) minor axis vs. area, (C) minor axis vs. major axis.



The estimated variance within lines of the inbred populations used in this study is shown in Figure 3. Measurements of variance can help in the design and analysis of experiments because they reveal aspects of the genetic attributes of the populations under study such as possibilities of linkage disequilibrium as well as the presence of additive or dominance effects (LYNCH 1988). Furthermore, the amount of phenotypic variability present within a line can be useful for plant breeders as it is one predictor of variability in lines derived from it (WHITLOCK and FOWLER 1996). Figure 3 displays the average trait value ± variance for each line of the RIL1 and RIL2 populations for seed area (Figure 3A), major axis (Figure 3B), and minor axis (Figure 3C). All lines had similar variances, and these variances were small compared to the differences in the average trait values, probably due to limited linkage disequilibrium and dominance effects of the alleles present in these lines.

When two plants intermate, the resulting progeny inherit nuclear DNA from both parents while the cytoplasm is inherited almost exclusively from the maternal parent. Cytoplasmic inheritance occurs when genes within the organelles of the cytoplasm, such as the chloroplasts and mitochondria, are passed down to the next generation. Occasionally, a phenotype can be solely due to the contribution of these cytoplasmic factors, as is the case for leaf color in *Mirabilis jilapa* (CORRENS 1909). Statistical analysis was used to determine if the maternal cytoplasmic inheritance affects the phenotypic variation in seed size traits. For the area trait, cytoplasm donor significantly affects the variation in seed area in the RIL1 population, but not in RIL2 (Figure 4A). A similar significant effect was seen in the RIL1 population for the minor axis trait (Figure 4C). Maternal cytoplasmic inheritance did not

Figure 3: Within-genotype variation of seed size traits. Average trait values for (A) area, (B) major axis, and (C) minor axis are plotted for the RIL1 (orange squares) and RIL2 (blue squares) populations. Bars represent ± variance.

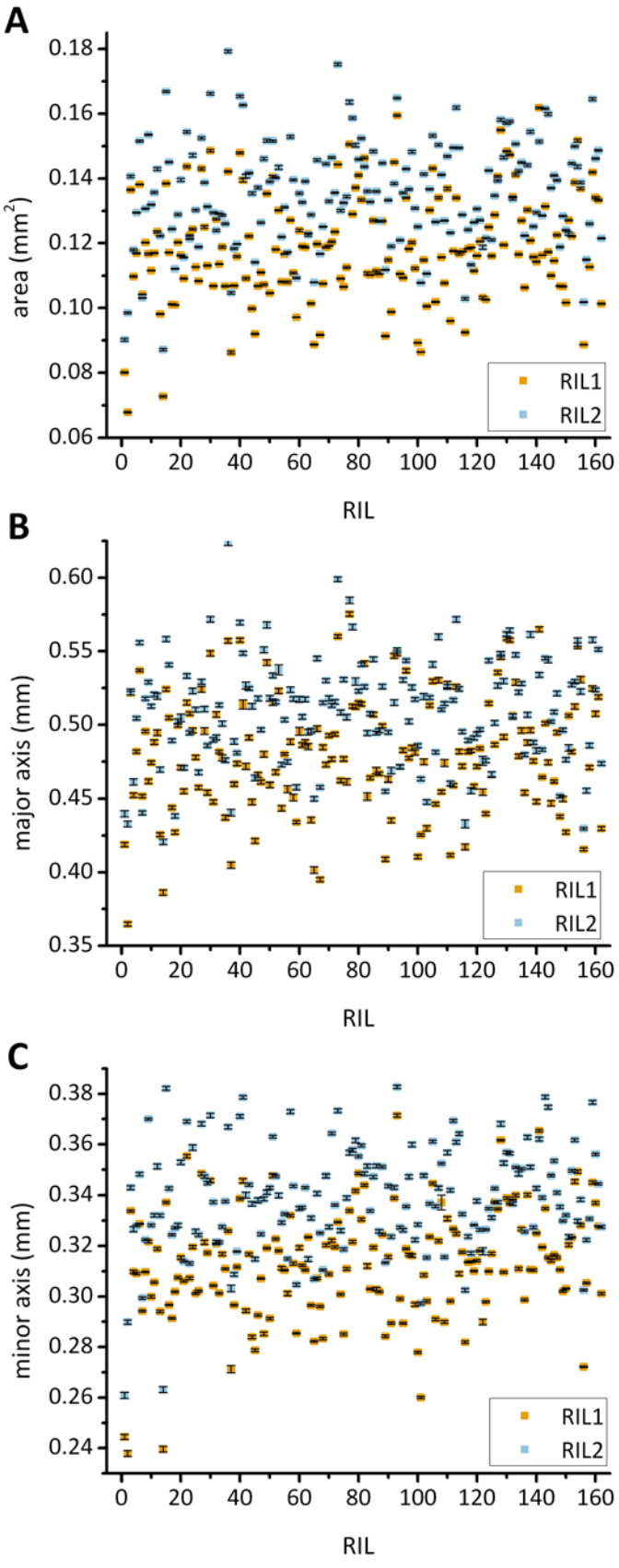
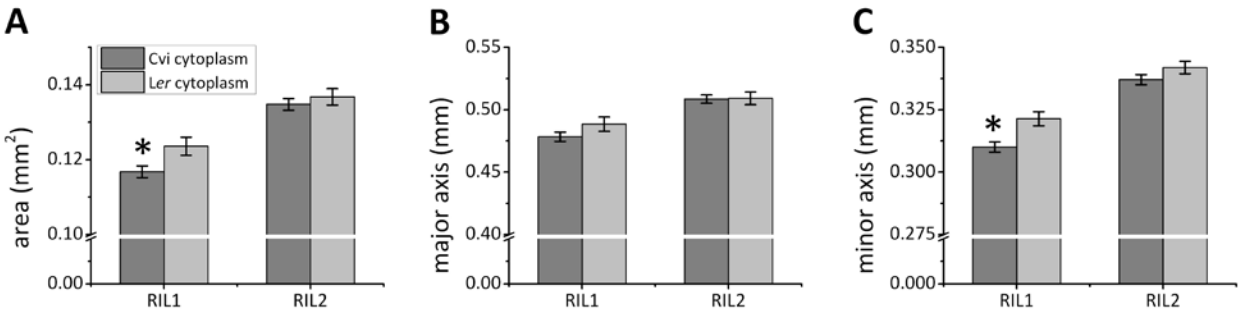

A
area (mm^2)
0.18
0.16
0.14
0.12
0.10
0.08
0.06
0
20
40
60
80
100
120
140
160
RIL
RIL1
RIL2
B
major axis (mm)
0.60
0.55
0.50
0.45
0.40
0.35
RIL
C
minor axis (mm)
0.38
0.36
0.34
0.32
0.30
0.28
0.26
0.24
RIL

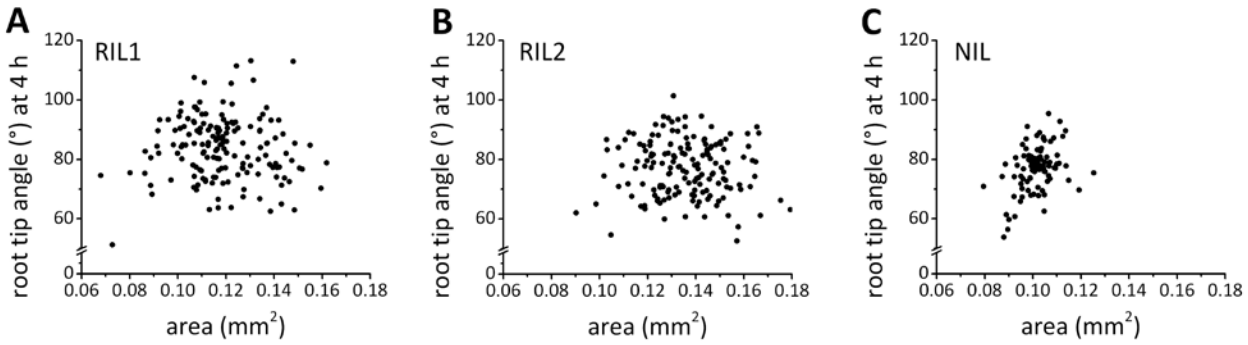
Figure 4: Effect of maternal cytoplasmic inheritance on seed size traits. The mean trait value ± SE is shown for lines carrying Cvi (dark gray bars) or L*er* (light gray bars) cytoplasm in both the RIL1 and RIL2 populations. Asterisks indicate significant differences between the two groups (α=0.05). (A) area, (B) major axis, (C) minor axis.



significantly affect the variation in major axis in either RIL population (Figure 4B). These results indicate the possibility that plastid DNA can be a contributing factor to seed size in Arabidopsis, and this involvement may be dependent on the environment in which the seeds were produced.

Experiments by Durham Brooks *et al.* (2010) measured the response to gravistimulation of Arabidopsis roots grown from one mother plant of the Columbia accession of Arabidopsis whose seeds were sieved into various seed size classes. In that study, the size of the seed was a major predictor of the resulting seedling's root gravitropic response. To determine if the seed sizes of lines from our inbred populations could predict their resulting gravitropism profile, the average root tip angle at 4 h post-gravistimulation was plotted against average seed area for lines of the RIL1 (Figure 5A), RIL2 (Figure 5B), or NIL (Figure 5C) populations. The 4 h time point was chosen because that is when the highest amount of variation between lines occurs. For both the RIL1 and RIL2 populations, no correlations between root tip angle and seed size could be determined. In the NIL population, a linear fit to the data showed a very weak correlation between the two traits (R^2 = 0.15). Because lines of the NIL population are composed mainly of L*er* DNA, with short introgressions of Cvi DNA, these data could support the hypothesis that seed size can help predict the parameters of root gravitropism when only comparing seeds of the same genotype, but not when considering multiple genotypes.

Figure 5: Correlations between seed size and root gravitropism. The mean root tip angle at 4 h post-gravistimulation for each genotype is plotted as a function of seed area for the (A) RIL1, (B) RIL2, and (C) NIL populations.



In Figure 6, statistical analysis was performed to determine if maternal cytoplasmic inheritance affects the development of root gravitropism. Student's t-tests were used to look for differences in root tip angle 0, 2, 4, and 8 h post-gravistimulation between seedlings with L*er* or Cvi cytoplasm. No significant differences were found at any time point for either the RIL1 (Figure 6A) or RIL2 (Figure 6B) populations. These data indicate that the genetic factors that determine variation in the root gravitropic response are most likely due only to nuclear DNA.

To capture as many root gravitropic responses as possible in the shortest amount of time, one 8 h experiment was begun in the morning (around 8 AM), and a second experiment was begun in the late afternoon (around 6 PM). These experiments took place on the lab bench top, not in a growth chamber, so the possibility exists that changes in building temperature from morning to evening could have an effect on the seedling's growth response. To calculate whether time of day had an effect on the experiments, the first and second principal components of tip angle development was calculated for each individual. Next, the average principal components of seedlings imaged in the morning were compared with those imaged in the evening for both the RIL1 and RIL2 populations. The individuals' data coordinates, as well as the average values ± the standard deviation are plotted in Figure 7. Student's t-tests found no significant differences in the root gravitropic response between those seedlings imaged in the morning and in the evening for either the RIL1 (Figure 7A) or RIL2 (Figure 7B) populations.

Another method to increase the phenotyping throughput of these experiments was to use multiple cameras at once. Seven cameras were used for RIL1 experiments, and 11 cameras

Figure 6: Effect of maternal cytoplasmic inheritance on root gravitropism. The mean trait values at 2 h intervals post-gravistimulation ± SE is shown for lines carrying Cvi (dark gray bars) or L*er* (light gray bars) cytoplasm in both the (A) RIL1 and (B) RIL2 populations. No significant differences were observed (α=0.05).

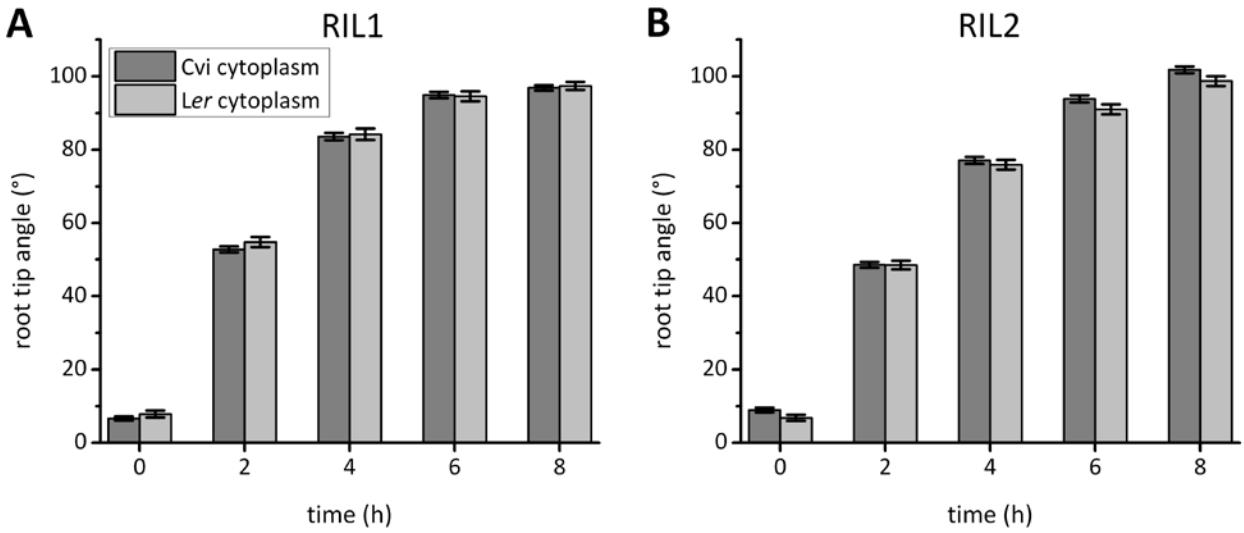
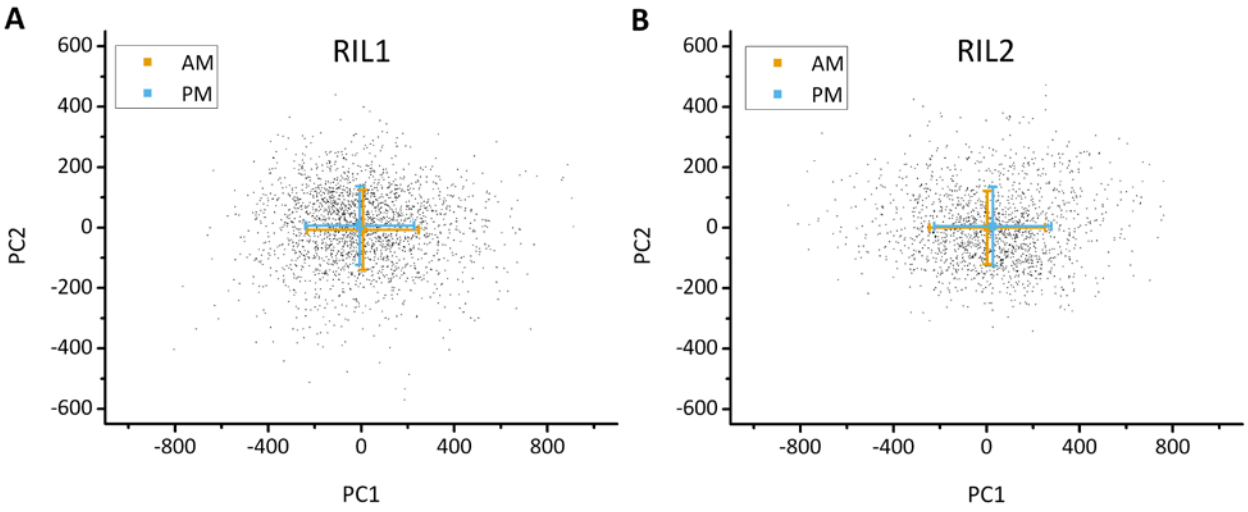


Figure 7: Effect of experiment time on root gravitropism. The first and second principal components of each root's tip angle development over time are plotted in small black circles. The average principal components of roots used in the AM (orange) and PM (blue) are displayed ± one standard deviation. (A) RIL1, (B) RIL2.



were used for RIL2 and NIL experiments. As in Figure 7, the first and second principal components of the root gravitropic response were calculated for each root, and then average values ± standard deviation were calculated for each camera. These data are plotted in Figure 8. Analysis of variance indicated a small but significant effect of camera position on the root gravitropic response for both the RIL1 (Figure 8A) and RIL2 (Figure 8B) experiments. These differences could be due to the position of air handlers in the room, proximity to the door, differences in light levels at each camera position, or other unknown factors. However, the effect of camera position on the root gravitropic response should not significantly alter our interpretation of the QTL analyses because great care was taken to randomize planting and positioning of the inbred lines.

Because the first and second principal components of tip angle development can explain most of the variation in the root gravitropic response, these values were used as phenotype data in a QTL analysis. The LOD profiles of significant QTL found via the standard stepwise model selection procedure are shown in Figure 9. For RIL1, four QTL were found using the first principal component as the phenotype value (1@60.0, 4@30.0, 4@38.0, and 5@51.9) while no significant loci were detected using the second principal component (Figure 9A,B). For RIL2, three QTL were significant for the first principal component (1@67.2, 3@15.1, and 4@36), and two QTL were identified for the second principal component (4@1.0 and 5@12.5) (Figure 9C,D). For NIL, two loci were found for both the first (3@9.0 and 3@42.5) and second (4@85.0 and 5@104.0) principal components (Figure 9E,F). The positions of QTL found when using the first principal component values as the phenotypes in the QTL analyses closely match the loci

Figure 8: Effect of camera position on root gravitropism. The first and second principal components of each root's tip angle development over time are plotted in small black circles. The average principal components of roots imaged with each camera are shown with color, with error bars indicating ± one standard deviation. (A) RIL1, (B) RIL2.

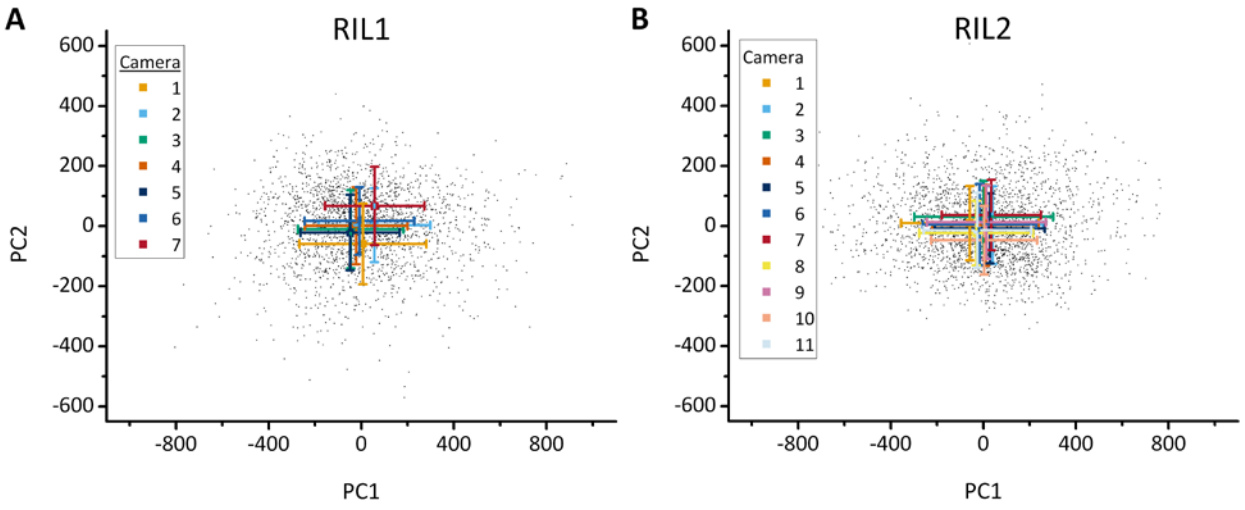
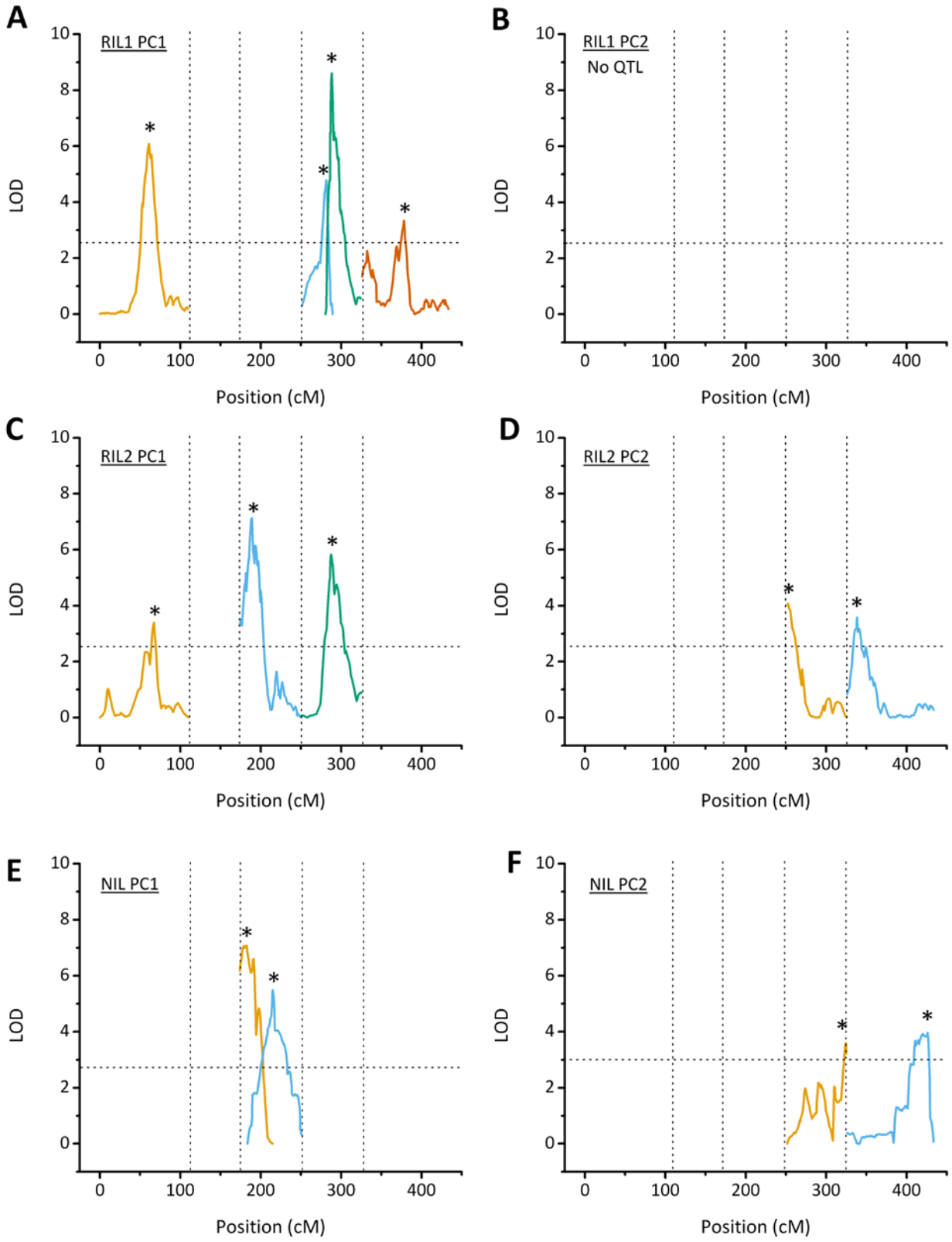


Figure 9: LOD profiles of the first and second principal components of the root gravitropic response statistically modeled as a function of genotype. Root gravitropism QTL were determined by multiple-interval mapping of the first (PC1) and second (PC2) principal components of the gravitropic response using the two independent RIL datasets (A–B, C–D) and the NIL dataset (E–F). (A) PC1 QTL of the RIL1 dataset, (B) PC2 QTL of the RIL1 dataset, (C) PC1 QTL of the RIL2 dataset, (D) PC2 QTL of the RIL2 dataset, (E) PC1 QTL of the NIL dataset, (F) PC2 QTL of the NIL dataset. Asterisks denote the position with the highest LOD score for each locus. Vertical dotted lines are used to separate the five chromosomes. Horizontal, dotted lines indicate the significance threshold (α=0.05).


A
RIL1 PC1
B
RIL1 PC2
No QTL
C
RIL2 PC1
D
RIL2 PC2
E
NIL PC1
F
NIL PC2
LOD
Position (cM)
0
2
4
6
8
10
0
100
200
300
400

identified using both stepwise QTL analysis at each time point and the SLOD method of locating QTL (Chapter 3). However, none of the loci identified using the second principal component as the phenotype value matched the loci from earlier analyses. Although using the principal components instead of raw tip angles results in similar QTL models, the principal component method does not provide information on the timing of action of these loci during the response. Additionally, estimating the effect of substituting one parental genome at a particular locus is no longer in units of degrees, obscuring the interpretation of how much substitution at that locus would change the tip angle at a certain time. Still, using the principal components as the phenotypes for QTL analyses is computationally less demanding than an analysis at each time point.

References

ABELSON, P. H., and P. J. HINES, 1999 The plant revolution. Science **285:** 367-368.

AHMADIYEH, N., G. A. CHURCHILL, K. SHIMOMURA, L. C. SOLBERG, J. S. TAKAHASHI *et al.*, 2003 X-linked and lineage-dependent inheritance of coping responses to stress. Mamm. Genome **14:** 748-757.

ALIMI, N. A., M. C. A. M. BINK, J. A. DIELEMAN, M. NICOLAI, M. WUBS *et al.*, 2013 Genetic and QTL analyses of yield and a set of physiological traits in pepper. Euphytica **190:** 181-201.

ALONSO-BLANCO, C., M. G. M. AARTS, L. BENTSINK, J. J. B. KEURENTJES, M. REYMOND *et al.*, 2009 What has natural variation taught us about plant development, physiology, and adaptation? Plant Cell **21:** 1877-1896.

ALONSO-BLANCO, C., H. BLANKESTIJN-DE VRIES, C. J. HANHART and M. KOORNNEEF, 1999 Natural allelic variation at seed size loci in relation to other life history traits of *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA **96:** 4710-4717.

ALONSO-BLANCO, C., A. J. M. PEETERS, M. KOORNNEEF, C. LISTER, C. DEAN *et al.*, 1998 Development of an AFLP based linkage map of L*er*, Col and Cvi *Arabidopsis thaliana* ecotypes and construction of a L*er*/Cvi recombinant inbred line population. Plant J. **14:** 259-271.

ARMENGAUD, P., K. ZAMBAUX, A. HILLS, R. SULPICE, R. J. PATTISON *et al.*, 2009 EZ-Rhizo: integrated software for the fast and accurate measurement of root system architecture. Plant J. **57:** 945-956.

ATWELL, S., Y. S. HUANG, B. J. VILHJALMSSON, G. WILLEMS, M. HORTON *et al.*, 2010 Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. Nature **465:** 627-631.

BALASUBRAMANIAN, S., C. SCHWARTZ, A. SINGH, N. WARTHMANN, M. C. KIM *et al.*, 2009 QTL mapping in new *Arabidopsis thaliana* advanced intercross-recombinant inbred lines. PLoS One **4:** e4318.

BALDWIN, K. L., A. K. STROHM and P. H. MASSON, 2013 Gravity sensing and signal transduction in vascular plant primary roots. Am. J. Bot. **100:** 126-142.

BASU, P., A. PAL, J. P. LYNCH and K. M. BROWN, 2007 A novel image-analysis technique for kinematic study of growth and curvature. Plant Physiol. **145:** 305-316.

BENGOUGH, A. G., D. C. GORDON, H. AL-MENAIE, R. P. ELLIS, D. ALLAN *et al.*, 2004 Gel observation chamber for rapid screening of root traits in cereal seedlings. Plant Soil **262:** 63-70.

BILLIAU, K., H. SPRENGER, C. SCHUDOMA, D. WALTHER and K. I. KÖHL, 2012 Data management pipeline for plant phenotyping in a multisite project. Funct. Plant Biol. **39:** 948-957.

BLANCAFLOR, E. B., and P. H. MASSON, 2003 Plant gravitropism. Unraveling the ups and downs of a complex process. Plant Physiol. **133:** 1677-1690.

BLUM, H., 1967 A transformation for extracting new descriptors of shape, pp. 362-380 in *Models for the Perception of Speech and Visual Form: Proceedings of a Symposium*, edited by W. WATHEN-DUNN. M.I.T. Press.

BOONSIRICHAI, K., C. GUAN, R. CHEN and P. H. MASSON, 2002 Root gravitropism: an experimental tool to investigate basic cellular and molecular processes underlying mechanosensing and signal transmission in plants. Annu. Rev. Plant Biol. **53:** 421-447.

BRACHI, B., N. FAURE, M. HORTON, E. FLAHAUW, A. VAZQUEZ *et al.*, 2010 Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. PLoS Genet. **6:** e1000940.

BROMAN, K. W., and S. SEN, 2009 *A guide to QTL mapping with R/qtl*. Springer, New York.

BROMAN, K. W., H. WU, S. SEN and G. A. CHURCHILL, 2003 R/qtl: QTL mapping in experimental crosses. Bioinformatics **19:** 889-890.

CALENGE, F., V. SALIBA-COLOMBANI, S. MAHIEU, O. LOUDET, F. DANIEL-VEDELE *et al.*, 2006 Natural variation for carbohydrate content in Arabidopsis. Interaction with complex traits dissected by quantitative genetics. Plant Physiol. **141:** 1630-1643.

CASPAR, T., and B. G. PICKARD, 1989 Gravitropism in a starchless mutant of *Arabidopsis*. Implications for the starch-statolith theory of gravity sensing. Planta **177:** 185-197.

CASSON, S. A., J. F. TOPPING and K. LINDSEY, 2009 MERISTEM-DEFECTIVE, an RS domain protein, is required for the correct meristem patterning and function in Arabidopsis. Plant J. **57:** 857-869.

CHAN, E. K. F., H. C. ROWE, J. A. CORWIN, B. JOSEPH and D. J. KLIEBENSTEIN, 2011 Combining genome-wide association mapping and transcriptional networks to identify novel genes controlling glucosinolates in *Arabidopsis thaliana*. PLoS Biol. **9:** e1001125.

CHAUDHURY, A. M., A. KOLTUNOW, T. PAYNE, M. LUO, M. R. TUCKER *et al.*, 2001 Control of early seed development. Annu. Rev. Cell Dev. Bi. **17:** 677-699.

CHAUDHURY, A. M., L. MING, C. MILLER, S. CRAIG, E. S. DENNIS *et al.*, 1997 Fertilization-independent seed development in *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA **94:** 4223-4228.

CHAVARRIA-KRAUSER, A., K. A. NAGEL, K. PALME, U. SCHURR, A. WALTER *et al.*, 2008 Spatio-temporal quantification of differential growth processes in root growth zones based on a novel combination of image sequence processing and refined concepts describing curvature production. New Phytol. **177:** 811-821.

CHEN, R., P. HILSON, J. SEDBROOK, E. ROSEN, T. CASPAR *et al.*, 1998 The *Arabidopsis thaliana AGRAVITROPIC 1* gene encodes a component of the polar-auxin-transport efflux carrier. Proc. Natl. Acad. Sci. USA **95:** 15112-15117.

CHURCHILL, G. A., and R. W. DOERGE, 1994 Empirical threshold values for quantitative trait mapping. Genetics **138:** 963-971.

CLARK, R. T., R. B. MACCURDY, J. K. JUNG, J. E. SHAFF, S. R. MCCOUCH *et al.*, 2011 Three-dimensional root phenotyping with a novel imaging and software platform. Plant Physiol. **156:** 455-465.

COLE, B., S. A. KAY and J. CHORY, 2011 Automated analysis of hypocotyl growth dynamics during shade avoidance in Arabidopsis. Plant J. **65:** 991-1000.

CORRENS, C., 1909 Vererbungsversuche mit blass (gelb) grünen und buntblättrigen Sippen bei *Mirabilis*, *Urtica* und *Lunaria*. ZIAV **1:** 291-329.

COTSAFTIS, O., D. PLETT, A. A. T. JOHNSON, H. WALIA, C. WILSON *et al.*, 2011 Root-specific transcript profiling of contrasting rice genotypes in response to salinity stress. Mol. Plant **4:** 25-41.

COUVARES, P., T. KOSAR, A. ROY, J. WEBER and K. WENGER, 2007 *Workflow in Condor*. Springer Press.

CROSS, J. M., M. VON KORFF, T. ALTMANN, L. BARTZETKO, R. SULPICE *et al.*, 2006 Variation of enzyme activities and metabolite levels in 24 Arabidopsis accessions growing in carbon-limited conditions. Plant Physiol. **142:** 1574-1588.

DARVASI, A., and M. SOLLER, 1995 Advanced intercross lines, an experimental population for fine genetic mapping. Genetics **141:** 1199-1207.

DE FOLTER, S., R. G. H. IMMINK, M. KIEFFER, L. PAŘENICOVÁ, S. R. HENZ *et al.*, 2005 Comprehensive interaction map of the Arabidopsis MADS box transcription factors. Plant Cell **17:** 1424-1433.

DE SOUZA, N., 2010 High-throughput phenotyping. Nat. Methods **7:** 36-36.

DOUGLAS, D. A., 1991 Clonal architecture of *Salix-setchelliana* (gravel bar willow) in Alaska. Can. J. Bot. **69:** 590-596.

DOWNIE, H., N. HOLDEN, W. OTTEN, A. J. SPIERS, T. A. VALENTINE *et al.*, 2012 Transparent soil for imaging the rhizosphere. PLoS One **7:** e44276.

DURHAM BROOKS, T. L., N. D. MILLER and E. P. SPALDING, 2010 Plasticity of Arabidopsis root gravitropism throughout a multidimensional condition space quantified by automated image analysis. Plant Physiol. **152:** 206-216.

ELWELL, A. L., D. S. GRONWALL, N. D. MILLER, E. P. SPALDING and T. L. BROOKS, 2011 Separating parental environment from seed size effects on next generation growth and development in Arabidopsis. Plant Cell Environ. **34:** 291-301.

ESAU, K., 1953 *Plant Anatomy*. John Wiley & Sons, Inc., New York.

ESHED, Y., and D. ZAMIR, 1995 An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. Genetics **141:** 1147-1162.

FABRE, J., M. DAUZAT, V. NEGRE, N. WUYTS, A. TIREAU *et al.*, 2011 PHENOPSIS DB: an information system for *Arabidopsis thaliana* phenotypic data in an environmental context. BMC Plant Biol. **11:** 77.

FALCONER, D. S., 1952 The problem of environment and selection. Amer. Nat. **86:** 293-298.

FALCONER, D. S., and T. F. C. MACKAY, 1996 *Introduction to Quantitative Genetics*. Longmans Green, Harlow, Essex, UK.

FANG, S., X. YAN and H. LIAO, 2009 3D reconstruction and dynamic modeling of root architecture in situ and its application to crop phosphorus research. Plant J. **60:** 1096-1108.

FANG, S. Q., R. T. CLARK, Y. ZHENG, A. S. IYER-PASCUZZI, J. S. WEITZ *et al.*, 2013 Genotypic recognition and spatial responses by rice roots. Proc. Natl. Acad. Sci. USA **110:** 2670-2675.

FILIAULT, D. L., and J. N. MALOOF, 2012 A genome-wide association study identifies variants underlying the *Arabidopsis thaliana* shade avoidance response. PLoS Genet. **8:** e1002589.

FITA, A., B. PICÓ, A. J. MONFORTE and F. NUEZ, 2008 Genetics of root system architecture using near-isogenic lines of melon. J. Am. Soc. Hortic. Sci. **133:** 448-458.

FITA, A., B. PICO, C. ROIG and F. NUEZ, 2007 Performance of *Cucumis melo* ssp. *agrestis* as a rootstock for melon. J. Hort. Sci. Biotechnol. **82:** 184-190.

FITA, A., A. RODRIGUEZ-BURRUEZO and J. PROHENS, 2010 Development and use of a new methacrylate container box for the study of root systems of horticultural plants. Bulletin of University of Agricultural Sciences and Veterinary Medicine Cluj-Napoca. Horticulture **67:** 245-247.

FITZ GERALD, J. N., M. D. LEHTI-SHIU, P. A. INGRAM, K. I. DEAK, T. BIESIADA *et al.*, 2006 Identification of quantitative trait loci that regulate Arabidopsis root system size and plasticity. Genetics **172:** 485-498.

FRENCH, A., S. UBEDA-TOMAS, T. J. HOLMAN, M. J. BENNETT and T. PRIDMORE, 2009 High-throughput quantification of root growth using a novel image-analysis tool. Plant Physiol. **150:** 1784-1795.

GALKOVSKYI, T., Y. MILEYKO, A. BUCKSCH, B. MOORE, O. SYMONOVA *et al.*, 2012 GiA Roots: software for the high throughput analysis of plant root system architecture. BMC Plant Biol. **12:** 116.

GRANIER, C., L. AGUIRREZABAL, K. CHENU, S. J. COOKSON, M. DAUZAT *et al.*, 2006 PHENOPSIS, an automated platform for reproducible phenotyping of plant responses to soil water deficit in *Arabidopsis thaliana* permitted the identification of an accession with low sensitivity to soil water deficit. New Phytol. **169:** 623-635.

GROSSNIKLAUS, U., J. P. VIELLE-CALZADA, M. A. HOEPPNER and W. B. GAGLIANO, 1998 Maternal control of embryogenesis by *MEDEA*, a Polycomb group gene in Arabidopsis. Science **280:** 446-450.

GRUBER, B. D., R. F. H. GIEHL, S. FRIEDEL and N. VON WIRÉN, 2013 Plasticity of the Arabidopsis root system under nutrient deficiencies. Plant Physiol. **163:** 161-179.

HALEY, C. S., and S. A. KNOTT, 1992 A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity **69:** 315-324.

HALEY, S. D., L. K. AFANADOR, P. N. MIKLAS, J. R. STAVELY and J. D. KELLY, 1994 Heterogeneous inbred populations are useful as sources of near-isogenic lines for RAPD marker localization. Theor. Appl. Genet. **88:** 337-342.

HASHIGUCHI, Y., M. TASAKA and M. T. MORITA, 2013 Mechanism of higher plant gravity sensing. Am. J. Bot. **100:** 91-100.

HERRIDGE, R. P., R. C. DAY, S. BALDWIN and R. C. MACKNIGHT, 2011 Rapid analysis of seed size in Arabidopsis for mutant and QTL discovery. Plant Methods **7:** 3.

HOBBS, D. H., J. E. FLINTHAM and M. J. HILLS, 2004 Genetic control of storage oil synthesis in seeds of Arabidopsis. Plant Physiol. **136:** 3341-3349.

HOSSAIN, S., R. FORD, D. MCNEIL, C. PITTOCK and J. F. PANOZZO, 2010 Development of a selection tool for seed shape and QTL analysis of seed shape with other morphological traits for selective breeding in chickpea (*Cicer arietinum* L.). Aust. J. Crop Sci. **4:** 278-288.

HUND, A., S. TRACHSEL and P. STAMP, 2009 Growth of axile and lateral roots of maize: I. Development of a phenotyping platform. Plant Soil **325:** 335-349.

INGRAM, P. A., J. ZHU, A. SHARIFF, I. W. DAVIS, P. N. BENFEY *et al.*, 2012 High-throughput imaging and analysis of root system architecture in *Brachypodium distachyon* under differential nutrient availability. Philos. T. Roy. Soc. B **367:** 1559-1569.

IYER-PASCUZZI, A. S., O. SYMONOVA, Y. MILEYKO, Y. HAO, H. BELCHER *et al.*, 2010 Imaging and analysis platform for automatic phenotyping and trait ranking of plant root systems. Plant Physiol. **152:** 1148-1157.

JIANG, C. J., and Z. B. ZENG, 1995 Multiple-trait analysis of genetic-mapping for quantitative trait loci. Genetics **140:** 1111-1127.

JOFUKU, K. D., B. G. W. DENBOER, M. VANMONTAGU and J. K. OKAMURO, 1994 Control of Arabidopsis flower and seed development by the homeotic gene *APETALA2*. Plant Cell **6:** 1211-1225.

JOFUKU, K. D., P. K. OMIDYAR, Z. GEE and J. K. OKAMURO, 2005 Control of seed mass and seed yield by the floral homeotic gene *APETALA2*. Proc. Natl. Acad. Sci. USA **102:** 3117-3122.

JOOSEN, R. V. L., D. ARENDS, L. A. J. WILLEMS, W. LIGTERINK, R. C. JANSEN *et al.*, 2012 Visualizing the genetic landscape of Arabidopsis seed performance. Plant Physiol. **158:** 570-589.

KAMOSHITA, A., J. WADE, L. ALI, S. PATHAN, J. ZHANG *et al.*, 2002 Mapping QTLs for root morphology of a rice population adapted to rainfed lowland conditions. Theor. Appl. Genet. **104:** 880-893.

KEARSEY, M. J., H. S. POONI and N. H. SYED, 2003 Genetics of quantitative traits in *Arabidopsis thaliana*. Heredity **91:** 456-464.

KEURENTJES, J. J., L. BENTSINK, C. ALONSO-BLANCO, C. J. HANHART, H. BLANKESTIJN-DE VRIES *et al.*, 2007 Development of a near-isogenic line population of *Arabidopsis thaliana* and comparison of mapping power with a recombinant inbred line population. Genetics **175:** 891-905.

KINOSHITA, T., R. YADEGARI, J. J. HARADA, R. B. GOLDBERG and R. L. FISCHER, 1999 Imprinting of the *MEDEA* polycomb gene in the Arabidopsis endosperm. Plant Cell **11:** 1945-1952.

KISS, J. Z., R. HERTEL and F. D. SACK, 1989 Amyloplasts are necessary for full gravitropic sensitivity in roots of *Arabidopsis thaliana*. Planta **177:** 198-206.

KIYOSUE, T., N. OHAD, R. YADEGARI, M. HANNON, J. DINNENY *et al.*, 1999 Control of fertilization-independent endosperm development by the *MEDEA* polycomb gene in Arabidopsis. Proc. Natl. Acad. Sci. USA **96:** 4186-4191.

KÖHLER, C., and U. GROSSNIKLAUS, 2002 Epigenetic inheritance of expression states in plant development: the role of Polycomb group proteins. Curr. Opin. Cell Biol. **14:** 773-779.

KÖHLER, C., L. HENNIG, C. SPILLANE, S. PIEN, W. GRUISSEM *et al.*, 2003 The Polycomb-group protein MEDEA regulates seed development by controlling expression of the MADS-box gene *PHERES1*. Gene Dev. **17:** 1540-1553.

KOORNNEEF, M., C. ALONSO-BLANCO and D. VREUGDENHIL, 2004 Naturally occurring genetic variation in *Arabidopsis thaliana*. Annu. Rev. Plant Biol. **55:** 141-172.

KOSAMBI, D. D., 1944 The estimation of map distances from recombination values. Ann. Eugenic **12:** 172-175.

KOVER, P. X., and R. MOTT, 2012 Mapping the genetic basis of ecologically and evolutionarily relevant traits in *Arabidopsis thaliana*. Curr. Opin. Plant Biol. **15:** 212-217.

KRANNITZ, P. G., L. W. AARSSEN and J. M. DOW, 1991 The effect of genetically based differences in seed size on seedling survival in *Arabidopsis thaliana* (Brassicaceae). Am. J. Bot. **78:** 446-450.

KUROMORI, T., T. WADA, A. KAMIYA, M. YUGUCHI, T. YOKOUCHI *et al.*, 2006 A trial of phenome analysis using 4000 Ds-insertional mutants in gene-coding regions of Arabidopsis. Plant J. **47:** 640-651.

LAFITTE, H. R., M. C. CHAMPOUX, G. MCLAREN and J. C. O'TOOLE, 2001 Rice root morphological traits are related to isozyme group and adaptation. Field Crop Res. **71:** 57-70.

LAM, L., S. W. LEE and C. Y. SUEN, 1992 Thinning methodologies - a comprehensive survey. IEEE T. Pattern Anal. **14:** 869-885.

LEMIEUX, B., M. MIQUEL, C. SOMERVILLE and J. BROWSE, 1990 Mutants of Arabidopsis with alterations in seed lipid fatty-acid composition. Theor. Appl. Genet. **80:** 234-240.

LEON-KLOOSTERZIEL, K. M., C. J. KEIJZER and M. KOORNNEEF, 1994 A seed shape mutant of Arabidopsis that is affected in integument development. Plant Cell **6:** 385-392.

LEWIS, D. R., N. D. MILLER, B. L. SPLITT, G. WU and E. P. SPALDING, 2007 Separating the roles of acropetal and basipetal auxin transport on gravitropism with mutations in two Arabidopsis *Multidrug Resistance-Like* ABC transporter genes. Plant Cell **19:** 1838-1850.

LI, Y.-F., G. KENNEDY, F. NGORAN, P. WU and J. HUNTER, 2013 An ontology-centric architecture for extensible scientific data management systems. Future Gener. Comp. Sy. **29:** 641-653.

LI, Z., P. MU, C. LI, H. ZHANG, Y. GAO *et al.*, 2005 QTL mapping of root traits in a doubled haploid population from a cross between upland and lowland japonica rice in three environments. Theor. Appl. Genet. **110:** 1244-1252.

LISEC, J., R. C. MEYER, M. STEINFATH, H. REDESTIG, M. BECHER *et al.*, 2008 Identification of metabolic and biomass QTL in *Arabidopsis thaliana* in a parallel analysis of RIL and IL populations. Plant J. **53:** 960-972.

LISTER, C., and C. DEAN, 1993 Recombinant inbred lines for mapping RFLP and phenotypic markers in *Arabidopsis thaliana*. Plant J. **4:** 745-750.

LIU, C. M., and D. W. MEINKE, 1998 The *titan* mutants of Arabidopsis are disrupted in mitosis and cell cycle control during seed development. Plant J. **16:** 21-31.

LOBET, G., L. PAGÈS and X. DRAYE, 2011 A novel image-analysis toolbox enabling quantitative analysis of root system architecture. Plant Physiol. **157:** 29-39.

LUO, M., P. BILODEAU, E. S. DENNIS, W. J. PEACOCK and A. CHAUDHURY, 2000 Expression and parent-of-origin effects for *FIS2*, *MEA*, and *FIE* in the endosperm and embryo of developing Arabidopsis seeds. Proc. Natl. Acad. Sci. USA **97:** 10637-10642.

LUO, M., E. S. DENNIS, F. BERGER, W. J. PEACOCK and A. CHAUDHURY, 2005 *MINISEED3* (*MINI3*), a *WRKY* family gene, and *HAIKU2* (*IKU2*), a leucine-rich repeat (*LRR*) *KINASE* gene, are regulators of seed size in Arabidopsis. Proc. Natl. Acad. Sci. USA **102:** 17531-17536.

LYNCH, M., 1988 Design and analysis of experiments on random drift and inbreeding depression. Genetics **120:** 791-807.

MANICHAIKUL, A., J. DUPUIS, S. SEN and K. W. BROMAN, 2006 Poor performance of bootstrap confidence intervals for the location of a quantitative trait locus. Genetics **174:** 481-489.

MANICHAIKUL, A., J. Y. MOON, S. SEN, B. S. YANDELL and K. W. BROMAN, 2009 A model selection approach for the identification of quantitative trait loci in experimental crosses, allowing epistasis. Genetics **181:** 1077-1086.

MANNING, P., K. HOUSTON and T. EVANS, 2009 Shifts in seed size across experimental nitrogen enrichment and plant density gradients. Basic Appl. Ecol. **10:** 300-308.

MARAGOS, P. A., and R. W. SCHAFER, 1986 Morphological skeleton representation and coding of binary images. IEEE T. Acoust. Speech **34:** 1228-1244.

MEYER, R. C., M. STEINFATH, J. LISEC, M. BECHER, H. WITUCKA-WALL *et al.*, 2007 The metabolic signature related to high plant growth rate in *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA **104:** 4759-4764.

MEYEROWITZ, E. M., 2001 Prehistory and history of Arabidopsis research. Plant Physiol. **125:** 15-19.

MILLAR, A. A., and L. KUNST, 1997 Very-long-chain fatty acid biosynthesis is controlled through the expression and specificity of the condensing enzyme. Plant J **12:** 121-131.

MILLAR, A. A., and L. KUNST, 1999 The natural genetic variation of the fatty-acyl composition of seed oils in different ecotypes of *Arabidopsis thaliana*. Phytochemistry **52:** 1029-1033.

MILLER, N. D., T. L. DURHAM BROOKS, A. H. ASSADI and E. P. SPALDING, 2010 Detection of a gravitropism phenotype in *glutamate receptor-like 3.3* mutants of *Arabidopsis thaliana* using machine vision and computation. Genetics **186:** 585-593.

MILLER, N. D., G. B. MONSHAUSEN, E. P. SPALDING and S. GILROY, 2009 Automated computer vision measurements of relative elemental growth rates of auxin transport mutant. *American Society of Plant Biology*, Hawaii.

MILLER, N. D., B. M. PARKS and E. P. SPALDING, 2007 Computer-vision analysis of seedling responses to light and gravity. Plant J. **52:** 374-381.

MIYAMOTO, N., E. STEUDLE, T. HIRASAWA and R. LAFITTE, 2001 Hydraulic conductivity of rice roots. J. Exp. Bot. **52:** 1835-1846.

MOORE, C. R., D. S. GRONWALL, N. D. MILLER and E. P. SPALDING, 2013 Mapping quantitative trait loci affecting *Arabidopsis thaliana* seed morphology features extracted computationally from images. G3: Genes|Genomes|Genetics **3:** 109-118.

MORITA, M. T., 2010 Directional gravity sensing in gravitropism. Annu. Rev. Plant Biol. **61:** 705-720.

MOUSSEAU, T. A., and C. W. FOX, 1998 The adaptive significance of maternal effects. Trends Ecol. Evol. **13:** 403-407.

MÜLLER, A., C. GUAN, L. GALWEILER, P. TANZLER, P. HUIJSER *et al.*, 1998 *AtPIN2* defines a locus of Arabidopsis for root gravitropism control. EMBO J. **17:** 6903-6911.

NAEEM, A., A. P. FRENCH, D. M. WELLS and T. P. PRIDMORE, 2011 High-throughput feature counting and measurement of roots. Bioinformatics **27:** 1337-1338.

NAGEL, K. A., A. PUTZ, F. GILMER, K. HEINZ, A. FISCHBACH *et al.*, 2012 GROWSCREEN-Rhizo is a novel phenotyping robot enabling simultaneous measurements of root and shoot growth for plants grown in soil-filled rhizotrons. Funct. Plant Biol. **39:** 891-904.

NEMRI, A., S. ATWELL, A. M. TARONE, Y. S. HUANG, K. ZHAO *et al.*, 2010 Genome-wide survey of Arabidopsis natural variation in downy mildew resistance using combined association and linkage mapping. Proc. Natl. Acad. Sci. USA **107:** 10302-10307.

NOH, B., A. BANDYOPADHYAY, W. A. PEER, E. P. SPALDING and A. S. MURPHY, 2003 Enhanced gravi- and phototropism in plant *mdr* mutants mislocalizing the auxin efflux protein PIN1. Nature **423:** 999-1002.

NOH, B., A. S. MURPHY and E. P. SPALDING, 2001 Multidrug resistance-like genes of Arabidopsis required for auxin transport and auxin-mediated development. Plant Cell **13:** 2441-2454.

NORMANLY, J., 2012 *High-Throughput Phenotyping in Plants*. Humana Press.

O'NEILL, C. M., S. GILL, D. HOBBS, C. MORGAN and I. BANCROFT, 2003 Natural variation for seed oil composition in *Arabidopsis thaliana*. Phytochemistry **64:** 1077-1090.

O'NEILL, C. M., C. MORGAN, C. HATTORI, M. BRENNAN, U. ROSAS *et al.*, 2012 Towards the genetic architecture of seed lipid biosynthesis and accumulation in *Arabidopsis thaliana*. Heredity **108:** 115-123.

ORR, H. A., 1998 Testing natural selection vs. genetic drift in phenotypic evolution using quantitative trait locus data. Genetics **149:** 2099-2104.

OTSU, N., 1979 A threshold selection method from gray-level histograms. IEEE T. Syst. Man Cyb. **9:** 62-66.

PATERSON, A. H., J. W. DEVERNA, B. LANINI and S. D. TANKSLEY, 1990 Fine mapping of quantitative trait loci using selected overlapping recombinant chromosomes, in an interspecies cross of tomato. Genetics **124:** 735-742.

PRINZENBERG, A. E., H. BARBIER, D. E. SALT, B. STICH and M. REYMOND, 2010 Relationships between growth, growth response to nutrient supply, and ion content using a recombinant inbred line population in Arabidopsis. Plant Physiol. **154:** 1361-1371.

R CORE TEAM, 2013 R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria.

RAY, D. K., N. RAMANKUTTY, N. D. MUELLER, P. C. WEST and J. A. FOLEY, 2012 Recent patterns of crop yield growth and stagnation. Nat. Commun. **3:** 1293.

REITER, R. S., J. G. K. WILLIAMS, K. A. FELDMANN, J. A. RAFALSKI, S. V. TINGEY *et al.*, 1992 Global and local genome mapping in *Arabidopsis thaliana* by using recombinant inbred lines and random amplified polymorphic DNAs. Proc. Natl. Acad. Sci. USA **89:** 1477-1481.

RHODES, D., P. J. RICH, D. G. BRUNK, G. C. JU, J. C. RHODES *et al.*, 1989 Development of two isogenic sweet corn hybrids differing for glycinebetaine content. Plant Physiol. **91:** 1112-1121.

ROSZAK, P., and C. KOHLER, 2011 Polycomb group proteins are required to couple seed coat initiation to fertilization. Proc. Natl. Acad. Sci. USA **108:** 20826-20831.

RŮŽIČKA, K., L. C. STRADER, A. BAILLY, H. YANG, J. BLAKESLEE *et al.*, 2010 Arabidopsis *PIS1* encodes the ABCG37 transporter of auxinic compounds including the auxin precursor indole-3-butyric acid. Proc. Natl. Acad. Sci. USA **107:** 10749-10753.

SANGSTER, T. A., N. SALATHIA, S. UNDURRAGA, R. MILO, K. SCHELLENBERG *et al.*, 2008 HSP90 affects the expression of genetic variation and developmental stability in quantitative traits. Proc. Natl. Acad. Sci. USA **105:** 2963-2968.

SCHNEIDER, C. A., W. S. RASBAND and K. W. ELICEIRI, 2012 NIH Image to ImageJ: 25 years of image analysis. Nat. Meth. **9:** 671-675.

SCHRUFF, M. C., M. SPIELMAN, S. TIWARI, S. ADAMS, N. FENBY *et al.*, 2006 The *AUXIN RESPONSE FACTOR 2* gene of Arabidopsis links auxin signalling, cell division, and the size of seeds and other organs. Development **133:** 251-261.

SEEVERS, P. M., J. M. DALY and F. F. CATEDRAL, 1971 Role of peroxidase isozymes in resistance to wheat stem rust disease. Plant Physiol. **48:** 353-&.

SEN, S., and G. A. CHURCHILL, 2001 A statistical framework for quantitative trait mapping. Genetics **159:** 371-387.

SHAH, S., Z. G. XIN and J. BROWSE, 1997 Overexpression of the *FAD3* desaturase gene in a mutant of Arabidopsis. Plant Physiol. **114:** 1533-1539.

SKIRYCZ, A., K. VANDENBROUCKE, P. CLAUW, K. MALEUX, B. DE MEYER *et al.*, 2011 Survival and growth of Arabidopsis plants given limited water are not equal. Nat. Biotechnol. **29:** 212-214.

SOLBERG, L. C., A. E. BAUM, N. AHMADIYEH, K. SHIMOMURA, R. H. LI *et al.*, 2004 Sex- and lineage-specific inheritance of depression-like behavior in the rat. Mamm. Genome **15:** 648-662.

SONKA, M., V. HLAVAC and R. BOYLE, 2008 *Image Processing, Analysis, and Machine Vision*. Thompson Learning, Toronto.

SOZZANI, R. and P. N. BENFEY, 2011 High-throughput phenotyping of multicellular organisms: finding the link between genotype and phenotype. Genome Biol. **12:** 219.

SPALDING, E. P., 2009 Computer Vision as a Tool to Study Plant Development, pp. 317-326 in *Plant Systems Biology*, edited by D. A. BELOSTOTSKY. Totowa: Humana Press Inc.

SPALDING, E. P., 2013 Diverting the downhill flow of auxin to steer growth during tropisms. Am. J. Bot. **100:** 203-214.

SPALDING, E. P., and N. D. MILLER, 2013 Image analysis is driving a renaissance in growth measurement. Curr. Op. Plant Biol. **16:** 100-104.

STANGA, J. P., K. BOONSIRICHAI, J. C. SEDBROOK, M. S. OTEGUI and P. H. MASSON, 2009 A role for the TOC complex in Arabidopsis root gravitropism. Plant Physiol. **149:** 1896-1905.

STEINFATH, M., T. GARTNER, J. LISEC, R. C. MEYER, T. ALTMANN *et al.*, 2010 Prediction of hybrid biomass in *Arabidopsis thaliana* by selected parental SNP and metabolic markers. Theor. Appl. Genet. **120:** 239-247.

STOKES, D., C. MORGAN, C. O'NEILL and I. BANCROFT, 2007 Evaluating the utility of *Arabidopsis thaliana* as a model for understanding heterosis in hybrid crops. Euphytica **156:** 157-171.

STRADER, L. C., and B. BARTEL, 2009 The Arabidopsis PLEIOTROPIC DRUG RESISTANCE8/ABCG36 ATP Binding Cassette transporter modulates sensitivity to the auxin precursor indole-3-butyric acid. Plant Cell **21:** 1992-2007.

STRADER, L. C., and B. BARTEL, 2011 Transport and metabolism of the endogenous auxin precursor indole-3-butyric acid. Mol. Plant **4:** 477-486.

SUBRAMANIAN, R., E. P. SPALDING and N. J. FERRIER, 2013 A high throughput robot system for machine vision based plant phenotype studies. Mach. Vision Appl. **24:** 619-636.

TANABATA, T., T. SHIBAYA, K. HORI, K. EBANA and M. YANO, 2012 SmartGrain: High-throughput phenotyping software for measuring seed shape through image analysis. Plant Physiol. **160:** 1871-1880.

THAIN, D., T. TANNENBAUM and M. LIVNY, 2005 Distributed computing in practice: the Condor experience. Concurr. Comp.-Pract. E. **17:** 323-356.

TISNÉ, S., Y. SERRAND, L. BACH, E. GILBAULT, R. BEN AMEUR *et al.*, 2013 Phenoscope: an automated large-scale phenotyping platform offering high spatial homogeneity. Plant J. **74:** 534-544.

TOPP, C. N., A. S. IYER-PASCUZZI, J. T. ANDERSON, C. R. LEE, P. R. ZUREK *et al.*, 2013 3D phenotyping and quantitative trait locus mapping identify core regions of the rice genome controlling root architecture. Proc. Natl. Acad. Sci. USA **110:** E1695-1704.

TÖRJÉK, O., R. C. MEYER, M. ZEHNSDORF, M. TELTOW, G. STROMPEN *et al.*, 2008 Construction and analysis of two reciprocal Arabidopsis introgression line populations. J. Hered. **99:** 396-406.

TUINSTRA, M. R., G. EJETA and P. B. GOLDSBROUGH, 1997 Heterogeneous inbred family (HIF) analysis: a method for developing near-isogenic lines that differ at quantitative trait loci. Theor. Appl. Genet. **95:** 1005-1011.

UFAZ, S., and G. GALILI, 2008 Improving the content of essential amino acids in crop plants: goals and opportunities. Plant Physiol. **147:** 954-961.

ULMASOV, T., G. HAGEN and T. J. GUILFOYLE, 1999 Activation and repression of transcription by auxin-response factors. Proc. Natl. Acad. Sci. USA **96:** 5844-5849.

VAN DEN BRULE, S., and C. C. SMART, 2002 The plant PDR family of ABC transporters. Planta **216:** 95-106.

VAN DER HEIJDEN, G., Y. SONG, G. HORGAN, G. POLDER, A. DIELEMAN *et al.*, 2012 SPICY: towards automated phenotyping of large pepper plants in the greenhouse. Funct. Plant Biol. **39:** 870-877.

VAN DER WEELE, C. M., H. S. JIANG, K. K. PALANIAPPAN, V. B. IVANOV, K. PALANIAPPAN *et al.*, 2003 A new algorithm for computational image analysis of deformable motion at high spatial and temporal resolution applied to root growth. Roughly uniform elongation in the meristem and also, after an abrupt acceleration, in the elongation zone. Plant Physiol. **132:** 1138-1148.

VANKADAVATH, R. N., A. J. HUSSAIN, R. BODANAPU, E. KHARSHIING, P. O. BASHA *et al.*, 2009 Computer aided data acquisition tool for high-throughput phenotyping of plant populations. Plant Methods **5:** 18.

VAUGHN, L. M., and P. H. MASSON, 2011 A QTL study for regions contributing to *Arabidopsis thaliana* root skewing on tilted surfaces. G3: Genes|Genomes|Genetics **1:** 105-115.

VERRIER, P. J., D. BIRD, B. BURLA, E. DASSA, C. FORESTIER *et al.*, 2008 Plant ABC proteins – a unified nomenclature and updated inventory. Trends Plant Sci. **13:** 151-159.

VOORRIPS, R. E., A. PALLOIX, J. A. DIELEMAN, M. C. A. M. BINK, E. HEUVELINK *et al.*, 2010 Crop growth models for the -omics era: the EU-SPICY project, pp. 315-321 in *Advances in Genetics and Breeding of Capsicum and Eggplant: proceedings of the XIVth EUCARPIA Meeting on genetics and breeding of Capsicum and Eggplant*, Valencia, Spain.

WALTER, A., H. SPIES, S. TERJUNG, R. KÜSTERS, N. KIRCHGEßNER *et al.*, 2002 Spatio-temporal dynamics of expansion growth in roots: automatic quantification of diurnal course and temperature response by digital image sequence processing. J. Exp. Bot. **53:** 689-698.

WANG, A., D. GARCIA, H. ZHANG, K. FENG, A. CHAUDHURY *et al.*, 2010 The VQ motif protein IKU1 regulates endosperm growth and seed size in Arabidopsis. Plant J. **63:** 670-679.

WANG, L., I. V. UILECAN, A. H. ASSADI, C. A. KOZMIK and E. P. SPALDING, 2009 HYPOTrace: image analysis software for measuring hypocotyl growth and shape demonstrated on Arabidopsis seedlings undergoing photomorphogenesis. Plant Physiol. **149:** 1632-1637.

WEIGEL, D., and R. MOTT, 2009 The 1001 Genomes Project for *Arabidopsis thaliana*. Genome Biol. **10:** 107.

WESTERN, T. L., J. BURN, W. L. TAN, D. J. SKINNER, L. MARTIN-MCCAFFREY *et al.*, 2001 Isolation and characterization of mutants defective in seed coat mucilage secretory cell development in Arabidopsis. Plant Physiol. **127:** 998-1011.

WHITLOCK, M. C., and K. FOWLER, 1996 The distribution among populations in phenotypic variance with inbreeding. Evolution **50:** 1919-1926.

WOO, H. H., B. R. JEONG, K. B. KOO, J. W. CHOI, A. M. HIRSCH *et al*., 2007 Modifying expression of closely related UDP-glycosyltransferases from pea and Arabidopsis results in altered root development and function. Physiol. Plantarum **130:** 250-260.

WU, G., and E. P. SPALDING, 2007 Separate functions for nuclear and cytoplasmic cryptochrome 1 during photomorphogenesis of Arabidopsis seedlings. Proc. Natl. Acad. Sci. USA **104:** 18813-18818.

YADEGARI, R., T. KINOSHITA, O. LOTAN, G. COHEN, A. KATZ *et al.*, 2000 Mutations in the *FIE* and *MEA* genes that encode interacting polycomb proteins cause parent-of-origin effects on seed development by distinct mechanisms. Plant Cell **12:** 2367-2381.

YANG, Z. Y., H. DOWNIE, E. ROZBICKI, L. X. DUPUY and M. P. MACDONALD, 2013 Light Sheet Tomography (LST) for in situ imaging of plant roots. Opt. Express **21:** 16239-16247.

YAZDANBAKHSH, N., and J. FISAHN, 2009 High throughput phenotyping of root growth dynamics, lateral root formation, root architecture and root hair development enabled by PlaRoM. Funct. Plant Biol. **36:** 938-946.

ZAMIR, D., 2013 Where have all the crop phenotypes gone? PLoS Biol. **11:** e1001595.

ZEEGERS, M., F. RIJSDIJK and P. SHAM, 2004 Adjusting for covariates in variance components QTL linkage analysis. Behav. Genet. **34:** 127-133.

ZENG, Z. B., J. J. LIU, L. F. STAM, C. H. KAO, J. M. MERCER *et al.*, 2000 Genetic architecture of a morphological shape difference between two Drosophila species. Genetics **154:** 299-310.

ZHANG, X., R. J. HAUSE, JR. and J. O. BOREVITZ, 2012 Natural genetic variation for growth and development revealed by high-throughput phenotyping in *Arabidopsis thaliana*. G3: Genes|Genomes|Genetics **2:** 29-34.

ZHENG, B. S., L. YANG, W. P. ZHANG, C. Z. MAO, Y. R. WU *et al.*, 2003 Mapping QTLs and candidate genes for rice root traits under different water-supply conditions and comparative analysis across three populations. Theor. Appl. Genet. **107:** 1505-1515.

ZHENG, X. H., N. D. MILLER, D. R. LEWIS, M. J. CHRISTIANS, K. H. LEE *et al.*, 2011 *AUXIN UP-REGULATED F-BOX PROTEIN1* regulates the cross talk between auxin transport and cytokinin signaling during plant root growth. Plant Physiol. **156:** 1878-1893.

ZHOU, Y., X. ZHANG, X. KANG, X. ZHAO and M. NI, 2009 SHORT HYPOCOTYL UNDER BLUE1 associates with *MINISEED3* and *HAIKU2* promoters in vivo to regulate Arabidopsis seed development. Plant Cell **21:** 106-117.

ZHU, J., S. M. MICKELSON, S. M. KAEPPLER and J. P. LYNCH, 2006 Detection of quantitative trait loci for seminal root traits in maize (*Zea mays* L.) seedlings grown under differential phosphorus levels. Theor. Appl. Genet. **113:** 1-10.

ZHU, J. M., S. M. KAEPPLER and J. P. LYNCH, 2005 Mapping of QTLs for lateral root branching and length in maize (*Zea mays* L.) under differential phosphorus supply. Theor. Appl. Genet. **111:** 688-695.