GENETIC DISSECTION OF INFLORESCENCE ARCHITECTURE IN *ORYZA SATIVA* USING HIGH DENSITY GENOMIC DATA AND NOVEL PHENOTYPING STRATEGIES

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Samuel Vincent Crowell

May 2015

GENETIC DISSECTION OF INFLORESCENCE ARCHITECTURE IN *ORYZA SATIVA* USING HIGH DENSITY GENOMIC DATA AND NOVEL PHENOTYPING STRATEGIES

Samuel Vincent Crowell, Ph. D.

Cornell University 2015

As the carriers of fruits and grains, inflorescences have been an important target of selection throughout domestication and modern breeding in many crop species. In Asian rice (*Oryza sativa*), one of the world's staple crops, the inflorescence is known as a panicle, due to complex lateral branching patterns. Rice breeders have had trouble measuring panicle phenotypes, limiting their ability to assess the importance of candidate genes from model systems in the context of natural variation and incorporate these findings into field-level crop improvement pipelines. This dissertation describes the development of a novel open-source phenotyping platform, PANorama, which uses skeletonization techniques to extract a range of length, width, and count phenotypes from images of rice panicles. High-resolution phenotyping of panicle traits complemented dense genetic marker data, increasing both the number and effect size of quantitative trait loci (QTL) that were detected for panicle traits within a recombinant inbred line mapping population. Comparisons among traits across a diverse panel of varieties grown in the field revealed that relationships between distinct components of panicle architecture are conserved across *O. sativa*, despite at least two origins of domestication. However, panicle traits showed distinct

distributions within and between the rice subpopulations, suggesting that genetic variation has been partitioned during geographic and ecological radiation of the species. Genome-wide association (GWAS) revealed unique genetic architecture for different types of traits, and identified a large number of trait-specific, small-effect QTL associated with natural variation in panicle development, several of which are near known candidate genes. Finally, while panicle traits occasionally shared associations with traditional agronomic phenotypes and yield components, a large portion of inflorescence associations were located in unique regions of the genome, suggesting that relationships between panicle architecture and agronomic performance in rice are subtle and highly multi-genic. Taken together, these results provide a rich assessment of panicle development in *Oryza sativa* and establish a methodological framework for breeders interested in optimizing inflorescence architecture in the context of yield.

BIOGRAPHICAL SKETCH

Samuel Crowell is from Martinsburg, West Virginia, and is the son of Stephen Crowell and Lora Brown Crowell. Sam received his Bachelors of Science in Biology with Honors, *summa cum laude*, from West Virginia University (WVU) in 2010. In his first undergraduate biology class at WVU, Sam developed an interest in research science while performing an experiment on Mendelian inheritance in fruit flies. By January of his freshman year, he was already volunteering in a molecular biology lab, studying cold-shock domain proteins in Arabidopsis under Professor Dale Karlson. Working under Professor Karlson, Sam cloned his first green-fluorescent protein constructs and performed particle bombardment experiments to visualize cellular expression patterns before the end of the summer. In a matter of months, Sam had become fascinated with mechanistic molecular biology.

Over the next two summers, Sam received competitive internships from the National Science Foundation (NSF) to study at two of the country's leading public plant biology research institutions: The Donald Danforth Plant Science Center in St. Louis, Missouri, and the Boyce Thompson Institute in Ithaca, New York. Over the course of these internships, Sam developed a strong interest in the application of basic plant biology in addressing real-world problems in plant breeding programs. During his internship with the Boyce Thompson Institute, he met Professor Susan McCouch, who would eventually become his graduate thesis advisor. Upon matriculating from WVU, Sam entered the Ph.D. program in Plant Biology at Cornell University. After winning

funding from the NSF Graduate Research Fellowship Program during his first year of study, Sam joined Professor McCouch's lab and began his project on panicle development in domesticated Asian rice.

Sam's interest in multidisciplinary science led to several interesting collaborations throughout his dissertation. In collaboration with Professor Christopher Barrett in the Dyson School of Applied Economics and Management at Cornell University, Professor McCouch and Sam coauthored a book chapter describing crop technologies to policy makers, Chapter 2 herein. Additionally, Sam formed a fruitful collaboration with a visiting researcher from Brazil, Professor Alexandre Falcão. In collaboration with Professor Falcão, Sam developed and released a novel phenotyping platform for measuring panicle traits from images, Chapters 3 and 4. Finally, during the third year of his graduate work, Sam traveled to and managed a field trial at the International Rice Research Institute in Los Baños, Philippines in collaboration with Drs. Abdelbagi Ismail and Glenn Gregorio. The data he collected there became the basis of Chapter 5.

To my family

ACKNOWLEDGMENTS

I would like to thank my graduate thesis advisor Professor Susan McCouch for her guidance, support, encouragement, and understanding throughout the course of my doctoral program. I truly appreciate that she never said "no" to my ideas, and allowed me to pursue interesting avenues of research, discussion, and collaboration outside the traditional realms of scientific research. Additionally, I would like to thank my committee members, Professor Michael Scanlon and Professor Rebecca Nelson, for their input and wisdom during the course of my graduate degree. I have benefited personally and professionally from the range of expertise that my committee brought to every meeting, and I truly enjoyed the conversations and interactions that we shared.

I am especially grateful to my dedicated collaborator and friend, Professor Alexandre Falcão, without whom a significant portion of this dissertation would not have been possible. I extend a special thanks to Dr. Abdelbagi Ismail and Dr. Glenn Gregorio, who sponsored and mentored me during my time abroad in the Philippines. I would like to thank the Cornell Graduate School, the Cornell Departments of Plant Biology and Plant Breeding, and the International Rice Research Institute (IRRI), as I benefitted greatly from the training, funding, and resources they provided. I am extremely grateful for the assistance that I received from Lovelie Shaine Olivio and my field crew throughout harvest season at IRRI. I would like to thank the National Science Foundation, which has funded my research and training as a scientist through

both my undergraduate and graduate degrees. I am also thankful for the conversation and assistance that I have received from all of my Cornell RiceLab-mates.

Personally, I would like to thank the following for their friendship and support throughout my time in graduate school: Dr. Delbert Abi-Abdallah, Elizabeth Brauer, Edward Cobb, Dr. Yann de Mey, Martijn Mos, Dr. Karl Niklas, and Diane Wang. I am extremely grateful to my parents, brother, and sister, who have been loving and supportive in every aspect of my personal and professional life. Finally, I would like to extend my greatest thanks to my partner Dr. Joseph LeCates, who has been loving and supportive throughout my entire graduate career.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Meeting expected demands for food, feed, fiber, and fuel crops in the face of burgeoning global population within the coming decades will require overcoming significant social, political, economic, and technological hurdles. Gains in agricultural productivity in the developing world over the past 50 years were achieved through investment in crop research and development (R&D), extension, and education of smallholder farmers. However, despite increases in overall productivity, yield gains in staple food crops have stagnated over the past two decades, when compared to the growth observed from 1961 to 1990 (Alston et al. 2010). Slowed growth in crop productivity has paralleled decreases in publicly funded agricultural research, and over half of all crop R&D in the developing world now comes from the private sector (Pardey et al., 2006). While privately funded research has delivered several incredibly successful crop technologies, including high yielding hybrid maize and genetically engineered maize and cotton (Brookes and Barfoot, 2010), many crops have not benefitted from R&D due to patent restrictions or trade secrets. This has resulted in a significant technology gap, which limits delivery of improved varieties to the estimated 2.3 billion smallholder varieties in Asia and Africa (Barrett et al., 2001).

In some cases, existing technologies or management systems can address yield disparities, if coupled with investments in extension and outreach. In other cases, new crop technologies are needed to reinvigorate on farm productivity and profitability. Public-private partnerships represent one way of leveraging technologies from the private sector through negotiation around

complex patent restrictions (Ferroni and Castle, 2011). However, plant researchers in the public sector can also contribute to issues surrounding food security by investing in open-source breeding programs and developing new tools targeted for translational research. Collaborations between publicly funded research institutions, such as universities, national research and extension programs, and/or a Consultative Group on International Agriculture (CGIAR) center, could facilitate consolidation of scientific resources. By working together, researchers at collaborative institutions can evaluate crops across many environments, ensure progress towards common goals, and close the scientific gaps that limit crop technologies from reaching the farmers and breeders in the developing world that need them the most.

As one of the world's most important grain crops, domesticated Asian rice (*Oryza sativa*) has been the target of agriculture research and investment in the public sector for over 60 years. The first major push in research occurred during the Green Revolution, in which breeders developed semidwarf varieties that outperformed traditional landraces when grown under improved fertilizer regimens in irrigated paddies (Khush, 2001; Hedden, 2003). Within the past 20 years, there has been an explosive growth in the molecular characterization of rice via development of genetic markers and sequencing (International Rice Genome Sequencing Project, 2005; Xu et al., 2011; Sakai et al., 2014; Schatz et al., 2014). It is now well documented that *O. sativa* is comprised of two major varietal groups (*Indica* and *Japonica*) that are reflective of at least two independent domestication processes (Second, 1982; Garris et al., 2005). Due to the inbreeding nature of the species and geographic radiation throughout domestication, *O. sativa* has partitioned into five distinct subpopulations (*indica, aus, tropical japonica, temperate japonica, aromatic/Group V*), which harbor unique polymorphisms and morphological characteristics

(Garris et al., 2005; McNally et al., 2009; Huang et al., 2011; Zhao et al., 2011). The level of genetic and phenotypic variation present within rice, its tractable genome size (roughly 380 Megabases) (International Rice Genome Sequencing Project, 2005), and genetic synteny with other grains (Bolot et al., 2009) make *O. sativa* an excellent system in which to study both basic and applied research questions.

After the model organism *Arabidopsis thaliana*, rice was the second plant species to have a high-quality deoxyribonucleic acid (DNA) sequence backbone constructed using bacterial artificial chromosome (BAC) to BAC sequencing (Goff, 2002; Yu et al., 2002; International Rice Genome Sequencing Project, 2005). The rice pseudomolecules have been annotated by an international group of experts (Sakai et al., 2013). In recent years, high-density, publicly available marker resources have been developed for numerous cultivars (Feltus et al., 2004; Huang et al., 2010; Xu et al., 2011; Zhao et al., 2011; Yonemaru et al., 2014; Zhao et al., 2015), and over 3,000 complete genomes were recently published (3,000 rice genomes project, 2014). Additionally, large collections of gene expression data generated using microarrays are available for various stages of development (Sato et al., 2012), and roughly 700 genes have been functionally characterized (Yamamoto et al., 2012). With the deluge of "big data" in the rice community, it is becoming increasingly important to develop tools that facilitate access to and interpretation of large amounts of information. One of the final hurdles that prevents breeders from leveraging the vast resources described above is the limited availability of phenotyping technologies. Without accurate measurements of plant architecture and performance, it is difficult to separate what portion of phenotypic variation is the result of genetic polymorphism versus environmental noise.

Despite well-documented relationships between yield and inflorescence architecture in rice, there is a dearth of phenotypic information surrounding this critical stage of development. The rice inflorescence is classified as a panicle, a conical raceme in which seeds are carried on long branches (Ikeda et al., 2004). Position of a seed within a branch determines grain-filling rate, and the size of panicles affects the number of grains a plant can carry. However, breeders have always characterized panicle architecture using the gross morphological measurement "panicle length," because measuring highly branched structures is both time consuming and difficult to do manually. As a result, there is a significant gap in understanding regarding the genetic relationships between panicle architecture and yield, and the extent to which functionally characterized candidate genes within rice contribute to natural variation in morphology and/or yield. Clarifying these relationships within rice may provide insight into other related species, especially the major grains, which share similar molecular and developmental pathways in inflorescence development (Zhang and Yuan, 2014).

Within this dissertation, I begin with a discussion in Chapter 2 on crop technologies that are expected to make a positive impact on sociopolitical stability in the developing world over the next decade. The chapter was developed in equal collaboration with Professor Susan McCouch and is targeted for readers with expertise outside of the plant science disciplines. As such, the chapter integrates a range of topics including: descriptions of different plant breeding techniques and crop technologies; policy and patent issues that limit access to scientific resources; an analysis of the differences between private and publicly research investments; open-access

licensing and open-source breeding approaches; and novel technologies expected to make an impact on smallholder farms.

In Chapters 3 and 4, I focus on development of a new crop technology targeted for use by plant scientists in the public sector, a phenotyping platform named PANorama. PANorama was designed to measure a range of length, width, and count measurements from images of rice panicles. In Chapter 3, I discuss the logical framework in which PANorama was developed, and highlight specific features and resources that were created to facilitate its uptake within the public plant breeding community. Chapter 4 provides a detailed discussion regarding the implementation of the PANorama pipeline, which was designed in collaboration with Professor Alexandre Falcão. I discuss key features of the software that Dr. Falcão and I incorporated into the pipeline after rigorous beta testing. I then use PANorama's phenotyping capabilities to identify novel regions of the genome associated with panicle traits using quantitative trait loci (QTL) mapping in a rice recombinant inbred line mapping population. I also highlight PANorama's applicability in other plant species, and demonstrate that quantitative variation exists for numerous traits across a small number of diverse rice varieties.

Finally, in Chapter 5, I assess panicle architecture in field conditions across a large population of landraces that are representative of three subpopulations in *Oryza sativa*. I dissect panicle morphology in *O. sativa*, and highlight that many relationships between different types of traits are conserved in both the *Indica* and *Japonica* subspecies despite independent origins of domestication. Using genome-wide association (GWAS), I identify novel regions of the genome associated with reproductive morphology and link several candidate genes to natural variation in

panicle traits. I also discuss the subtle relationships between panicle architecture and yield performance in rice.

CHAPTER 2

CROP TECHNOLOGIES FOR THE COMING DECADE

Modified with permission from:

Susan McCouch and Samuel Crowell. (2013). Crop Technologies for the Coming Decade. In *Food Security and Sociopolitical Stability*, ed. Christopher B. Barrett. Oxford University Press.

Abstract. Our ability to meet future demand for food, feed, fiber, and fuel crops is challenged by rapid population and income growth, competition for scarce land and water resources, climate change, and major shifts in the funding and focus of global agricultural research. Innovations in crop technologies targeting the needs of smallholder farmers are key to ensuring food security in Asia and Africa, where demand is outpacing increases in productivity gains. However, agricultural research is becoming increasingly embedded in new layers of intellectual property protection, restrictions on the exchange of germplasm, and costly biosafety regulations that favor innovation by a well-endowed and profit-minded private sector. A reinvigoration of public-goods oriented agricultural research and extension is urgently needed to address this growing technology gap, which severely limits crop productivity for smallholder farmers in the developing world.

2.1 Introduction

Improvements in agricultural productivity over the last half-century have largely been achieved through technological change enabled by investments in agricultural research, development, extension, and education. While the world's population increased from 3.1 billion in 1960 to over seven billion in 2012, total cereal production grew faster than the population, from 643 million metric tons in 1960 to more than 2,280 million metric tons in 2010 (FAO 2010). This

boost in global crop productivity was brought about by a combination of crop genetics and breeding, drastically altered agronomic practices, greater access to credit and information by farmers, and enhanced real-time communication networks. These developments were supported by investments in agricultural research and development in both the public and private sectors.

Despite significant global increases in aggregate output since 1961, yield gains in staple food crops have slowed during 1990 to 2007 compared to 1961 to 1990 (Alston et al. 2010). With population and income growing rapidly around the world and climate change making extreme weather events more frequent, slowing yields raise serious concerns about our ability to sustain the increases in agricultural supply needed to meet future demands for food, feed, fiber, and fuel. Demand is growing fastest in Asia and Africa, where there is little spare land or water available for agricultural expansion. Thus, increasing crop productivity through technological change represents the best hope for improving food security in these regions.

Although large-holder farmers are crucial for maintaining affordable food supplies in urban centers, an estimated 2.3 billion people (one third of humankind) currently depends on income derived from small farms of less than two hectares in Asia and Africa (Barrett et al., 2001). Farm yields in key crops still vary significantly between farming regions, however, and the cost-benefit ratios of different agricultural practices often remain far below their potential, which results in a significant "yield and sustainability gap" (Byerlee et al., 2009). There are many reasons for the shortfalls, including fundamental resource constraints (poverty); insufficient information or education about improved agronomic practices; lack of timely access to inputs, such as fertilizer, pesticides, herbicides, or water; use of crop varieties that are susceptible to

pests, diseases, and abiotic stresses; postharvest losses due to poor storage facilities; and ineffective market integration. Existing technologies and management systems can sometimes address these disparities, when coupled with investments in extension and outreach. In other cases, new crop technologies are key ingredients that can boost the productivity and profitability of a farm operation.

2.2 Scientific haves and have-nots

Over the past two decades, publicly-funded agricultural R&D programs have been drastically reduced, while the private agricultural sector has experienced unprecedented growth—a trend that is particularly noticeable in wealthy countries since the 1990s (Pardey et al., 2006). This shift has brought about a suite of changes that affect every aspect of agriculture in both the developed and developing world. Mechanical and chemical technologies that generate greater returns on investments have become R&D's major focus, while development of new inbred varieties in staple food crops and environmental management strategies that enhance on-farm productivity has dwindled. With a strengthening of intellectual property (IP) rights, restrictions on the exchange of germplasm, introduction of costly biosafety regulations, and a shift toward endeavors that are more likely to maximize profit, such as improving food quality and developing novel medical, energy, and industrial applications of agricultural products (Pardey and Pingali, 2010), basic research and extension focused on increasing crop productivity has been critically underfunded.

As wealthy countries reorient their agricultural R&D away from the types of technologies that are most easily adopted by farmers in developing countries, they limit low-income producers'

rapid uptake of new technology. A significant geopolitical divide in ownership of scientific technology worldwide is leading to a growing gap between the scientific haves and have-nots (Pardey et al., 2006). Gene grabs, land grabs, and water wars undertaken by governments and private-sector institutions seeking to control scarce resources are manifestations of this divide, threatening political stability as poorer or weaker nations and populations are outcompeted for access to essential natural resources. Obstacles that continue to hinder the flow of improved germplasm, technology, and information deepen the gap between rich and poor, leading to greater vulnerability in resource-poor populations threatened by climate change and global economic instability.

Public-private partnerships that are explicitly pro-poor and have a commitment to R&D intended to bridge the technology gap and address the needs of smallholder farmers offer some hope of addressing the root causes of social unrest and political upheaval. Where the public sector typically lacks resources to make these investments—and the private sector often lacks the motivation to do so—foundations can step in and foster these partnerships. Complex institutional arrangements are often required to negotiate access to technology, establish biosafety protocols, and engage farmers and farming communities in their own development. If well designed, public-private partnerships can maximize gains for both investors and small-scale farmers (Ferroni and Castle, 2011).

2.3 Genetics and agronomics: Two sides of the same coin

We use the term "crop technologies" to refer to the combination of genetics (new crop varieties) and agronomic practices (management of soil, water, fertilizer, pests, diseases, and so on). Crop

technologies are multicomponent systems developed by multiple individuals or institutions, and they continue to evolve as farmers deploy them. Successful execution of crop technologies depends on farmers' ability to adapt basic ideas and innovations to new situations using locally available materials to integrate different components in different combinations —the old and the new, the low tech and the high tech. Ingenious combinations that optimize on-farm profitability, productivity, and input use efficiency are often developed by resource-poor farmers who need them the most, sometimes with the help of extension agents, nonprofit institutions, or private-sector agricultural advisers. Involving farmers and community members in development and dissemination of new crop technologies helps investments take root and invites local innovation that is needed to drive growth.

Over the last 50 to 60 years, yield gains in the world's major staple crops were achieved through a combination of intensive plant breeding and large-scale environmental engineering. Farmers optimized the performance of their cropping systems by combining improved seed, new forms of mechanization, and intensive management of water, fertilizer, pesticide, and herbicide—doubling or tripling yields per unit area. From the 1960s to the 1980s, the Green Revolution increased agricultural production around the world by fitting into this paradigm: international crop improvement efforts led by the International Maize and Wheat Center (CIMMYT) and the International Rice Research Institute (IRRI) provided high-yielding, semi-dwarf varieties of wheat and rice to farmers (with no intellectual property protection) and dramatically increased harvestable yields, but only if farmers could supply irrigation, fertilizer, and pest and disease management. A supportive policy environment also provided access to credit and expanded markets, both critical in driving the productivity gains made possible by improved genetics.

Today, this yield optimization paradigm has given way to new realities, with more attention to renewable resource management and long-term environmental impact. Slowing productivity gains, shifting global climate trends, and public outcry against contamination of food, soil, and waterways due to excessive use of agricultural inputs highlight the need for a fresh approach to variety development, pest management, and soil and water resource conservation.

Conservation agriculture evolved as a way for smallholder farmers to enhance soil resources and manage expensive chemical and water inputs (Hobbs, 2007), but these practices quickly permeated all levels of agriculture, as they represent a sustainable and economic way of managing farm resources at any level of income or production. Conservation agriculture involves three interlinked principles: (1) minimal mechanical soil disturbance (no-tillage), (2) permanent organic soil cover (living cover crop or green mulch), and (3) diversified crop rotations (FAO 2012). Zero tilling maintains a permanent or semipermanent organic soil cover, which physically protects the soil from sun, rain, and wind. Soil microorganisms and fauna are encouraged to perform functions, such as aeration of soil and nutrient release, which tillage used to play in conventional farming systems. A varied crop rotation scheme involving more than two crops enhances the diversity of the system both above and below ground, reducing disease and pest problems.

Conservation agriculture represents one approach towards developing more sustainable farming systems. However, there are still many questions concerning how basic principles and practices should be adapted and integrated into local farm conditions. Plant breeders are working to develop new varieties that fit into more sustainable cropping systems. They face questions about

which combinations of crops will be most productive, which traits are most urgently needed, what breeding technologies are most appropriate, and how to breed for a future when pests, diseases, and climate are extremely difficult to predict.

2.4 Access to plant genetic resources

Plant breeding's success depends on a breeder's ability to access, utilize, and exchange natural forms of genetic variation found in wild and domesticated plant species. Traditionally, farmers selectively maintained favorable plants and eliminated weak, diseased, or unattractive plants from wild or early-domesticated populations. Over time, this form of breeding led to the development of landraces of crops (or farmer varieties) and ornamental plants that provided the basis for virtually all of modern plant breeding. These early plant varieties and other genetic resources were historically considered part of the global genetic commons and could be freely collected and shared across the world. This notion of a shared genetic heritage led to the establishment of large botanical collections and the wide dissemination of valued crop plants. During the 20th century, the Consultative Group on International Agriculture (CGIAR) and the United Nations Food and Agriculture Organization (FAO) established national and international seed banks to collect, maintain, and disseminate genetically diverse strains of economically important plant species.

Ownership of genetic resources is now a major point of discussion affecting both public and private sector breeding programs. Traditional forms of collecting and exchanging germplasm are giving way to new realities. Intellectual property rights are being asserted over plant varieties at a rapid pace and countries are closing their borders to exchange of indigenous genetic resources.

The global genetic commons is shrinking at a very rapid pace, heightening tensions about access to both genetic resources and the technology needed to harness their potential—and deepening the divide between scientific haves and have-nots.

Plant Variety Protection and utility patents: Closing the commons

Treating the products of agricultural research as private goods is not a new concept. The U.S. Plant Patent Act (1930) allowed patent protection for asexually reproduced plants, focusing mostly on horticultural and ornamental species and specifically excluding potato. A second wave of appropriation started in Europe during the 1940s and '50s with the adoption of a system of plant breeders' rights. These rights were revised and extended to other countries with the establishment of the International Convention for the Protection of New Varieties of Plants (UPOV) in 1961. The system granted plant breeders a form of intellectual property rights (IPR) referred to as Plant Variety Protection (PVP) that is currently in place in about 70 countries. PVP allows plant breeders to protect varieties that they develop from appropriation by others, but it guarantees the breeders' exemption, allowing other breeders to use PVP-protected varieties to generate further varietal improvement. It also guarantees farmers' privilege, allowing farmers to save seed of PVP varieties for their own reuse. These provisions are designed to protect plant breeders' rights while ensuring continued access to plant genetic resources and sharing of benefits derived from their use.

While in many parts of the world crop genetic resources are still managed as public goods, the situation changed dramatically in the United States with the 1980 Diamond v. Chakrabarty Supreme Court decision. This decision allowed utility patents to be issued on plant varieties for

the first time. Utility patents confer the patent holder exclusive rights to exploit the "invention" and to exclude others from making, reproducing, using, selling, or importing the patented variety for a period of 20 years. There is neither breeders' exemption nor farmers' privilege, and unlike industrial utility patents, there is no requirement for the plant variety patent holder to disclose the technical process by which a new plant variety was developed. This contributes to a "closing of the commons," because the private innovator's rights supplant the rights of both the public breeding and traditional farming sectors, preventing access to technological innovation and enhanced germplasm that are the basis of innovation. Today, plants are eligible for utility patents in the U.S., Australia, and the EU, although in the EU a plant variety must be documented to carry a "particular gene" to be eligible for patent protection.

The 1995 Agreement on Trade-Related Aspects of Intellectual Property (TRIPS) inextricably ties trade with patent protection and is a requirement for members of the World Trade Organization (WTO) (Pardey et al., 2004). Countries participating in TRIPS are required to provide a system for IPR on plant varieties, involving either the use of patents or PVP as implemented under UPOV. Alternatively, a *sui generis* system of plant variety protection may be implemented, as in the 2001 Indian Protection of Plant Varieties and Farmers' Rights Act; it contains provisions for "benefit sharing," whereby local communities are acknowledged as contributors of landraces and farmer varieties in the breeding of new plant varieties. The Philippines and Thailand also implement *sui generis* protection systems. In some cases, countries have gotten together to develop regional IP regimes, such as the Andean Community (Bolivia, Colombia, Ecuador, Peru, and Venezuela), which expressly prohibits patents on plants and animals.

The Convention on Biological Diversity: Redefining the commons

In 1993, a decade after Diamond v. Chakrabarty, the United Nations ratified the Convention on Biological Diversity (CBD), which states that a country has sovereign rights to its indigenous genetic resources. The CBD sought to establish a mechanism by which countries could assign value to genetic resources and derive benefit from making them available as inputs for breeding new varieties and other forms of agricultural R&D, institutionalizing a dramatic change in the concept of ownership in the global genetic commons.

Most countries interpreted the CBD to mean that they may require payment for access to germplasm. Subsequently, the 2004 International Treaty on Plant Genetic Resources for Food and Agriculture was developed to determine how benefits created using these resources would be equitably shared and establish a multilateral system to facilitate access to genetic resources. The treaty currently governs the international exchange of germplasm for 29 species of forage crops and 35 species of food crops, including major the grains (rice, wheat, and maize). However, it excludes several major crops, including soybean and peanut, and has no provisions for biofuel crops. It is also vague on a number of points, including how monetary benefits from commercial products developed from gene bank materials are to flow. Farmers who conserve indigenous genetic resources are supposed to benefit, especially those in developing and transitional economies, but the particulars of how this will happen have yet to be clarified.

The lack of clarity in the CBD has taken a toll on international germplasm exchange and has created an environment in which commercial breeders often look for a work-around to access genetic materials that were distributed prior to 1993 when the CBD went into effect (Hammond,

2010). For example, U.S. gene banks maintain a major portion of the materials from the international plant germplasm collections funded by the Consultative Group on International Agricultural Research (CGIAR). Most of these materials were freely shared prior to 1993, and because the U.S. has signed but not ratified the treaty, it allows these materials to be accessed without applying the provisions of the treaty; that is, without the mandate for sharing the benefits of commercialization stipulated by the CBD. Many of these materials are now being conserved in private germplasm collections with no requirement to distribute them further. It is ironic that the CBD, aiming for access and benefit sharing, appears for the moment to have accomplished the opposite (Hammond, 2010).

Paying for gene banks: Who owns the commons?

Following Diamond v. Chakrabarty, commercial breeders and some governments began to amass large collections of improved germplasm for their own use. Despite the fact that many of these resources come directly from the public sector, genetic materials maintained in private collections are generally not shared. They guarantee the owners access to the large reservoirs of genetic variation needed to breed and patent new varieties. Yet private sector breeders still look to the public sector to collect, characterize, and perform prebreeding enhancement of landrace and wild relatives (Duvick, 1991). This dichotomy in the way germplasm resources are valued and paid for lies at the root of an uneasy relationship between public and private plant breeders today. Varieties developed by the private sector are protected by formal IPR, in contrast to wild relatives and landraces that are still considered a public good (Kronstad, 1996).

Private breeding enterprises and seed companies benefit greatly from access to publicly available genetic resources, but have so far been unwilling to contribute financially to long-term international germplasm conservation efforts. Rather, the strategy seems to be for major multinational breeding companies and some governments to try to appropriate entire collections of germplasm from international organizations such as the International Rice Research Institute or the International Maize and Wheat Improvement Center, or other countries that still support free exchange of germplasm. This gene grab or genetics arms race is taking place across the globe. As the strategic importance of germplasm diversity is obvious, the closure of the commons creates urgency in the face of potential scarcity.

Many breeding systems in the developing world are poorly equipped to deal with the rapid changes that are occurring, because they have depended so extensively on international free exchange of germplasm. As plant research in the industrialized world has come to be dominated by private companies, the uptake of new technologies and the spillover effects of global investments in agricultural R&D have slowed. Additionally, there is currently no mechanism for ensuring that the benefits of modern breeding will reach breeders and farmers in the developing world. Many argue that farmers in developing countries made an essential contribution to plant breeding and modern crop variety development by developing landraces, but received no payment or recognition for that contribution (Day, 1997). The financial incentives and rewards of the patent system that have been put in place to drive innovation in modern plant breeding are often in conflict with the essential "public goods" nature of the germplasm resources and crop varieties that underpin those innovations. New mechanisms for funding international, public-sector agricultural R&D are urgently needed to ensure that the benefits of modern plant breeding

translate into appropriate products and information systems that are made available to those in need of them in the developing world.

2.5 Classical plant breeding technology

Plant breeding has been practiced for thousands of years—in many cases inadvertently. Domestication of landraces from wild relatives represents the simplest form of plant breeding: selection of desirable individuals from existing variation, with no conscious crossing (hybridization) to generate new combinations of genes. Classical breeding exerts selection on diverse populations of plants in similar ways to nature, by imposing non-random forms of mating between closely or distantly related individuals (cross-pollination) and allowing some plants to reproduce at the expense of others every generation. Humans typically impose more intense selection pressure than nature does, speeding the process. Sexual crosses underpin the entire business of plant breeding in both the public and private sectors and are the foundation upon which all other breeding technologies are built. Newer approaches simply increase the efficiency, precision, or profitability of classical plant breeding—none has yet replaced it.

The level of technical knowledge required to manage a classical breeding program is usually relatively low, but a high level of decision-making and management expertise is required to coordinate a successful plant improvement program. It takes years of experience to become familiar with a crop, its germplasm, and the range of environments in which it is grown. A breeder may make and evaluate dozens to thousands of crosses per year, and these crosses may be evaluated by teams of professional breeders or done in on-farm trials in consultation with farmers. In each generation the breeder selects the lines with the most desirable traits and

continues the breeding cycle. This cyclical process recombines and recycles genes that, for the most part, have been available in the cultivated gene pool for thousands of years. It is important to note that there is no requirement to describe the underlying genetic composition of a new variety that has been bred using classical crossing and selection. In most cases, the exact genetic make-up is not known.

The time it takes from the first cross to final release of a new variety is generally between 10 and 15 years for annual crop species and may be considerable longer for perennials. Breeding populations must be evaluated in different years and locations to provide insight into performance stability in different target environments; after identification, several more years will pass before the release of successful varieties. With few exceptions, the public sector breeds self-pollinating, inbred species at government-supported National Agricultural Research and Extension Systems (NARES) and universities. Because self-pollinators are genetically uniform and perform stably when seed is saved from one generation to the next, and are thus difficult to protect with IPR, the private sector has little interest in inbreds. International crop improvement centers have and continue to play a major role in supporting breeding activities in inbreeding crops by collaborating with NARES partners—tailoring the majority of varieties that currently meet smallholder needs.

2.6 Hybrid breeding technology

Hybrid varieties were the primary driver of private investment in the plant breeding and seed business well before the 1990s (Shull, 1909). Hybrid breeding began in 1909 when Shull and colleagues demonstrated that by carefully mating genetically distinct, inbred parents, vigorous

corn hybrids could be produced that are genetically uniform and show superior performance to either parent. Parental lines are maintained by inbreeding and their identity is protected as a 'trade secret', while the offspring generated from crossing parental lines (F1 hybrid seed) is generated annually and sold every season to farmers. The genetic mechanisms controlling this phenomenon, known as heterosis or hybrid vigor, are still not entirely understood today.

The time required to develop a new hybrid variety is also 10 to 15 years. In terms of breeding, however, hybrid varieties are considerably more complicated and expensive to develop, due to integration of elaborate male sterility systems and rigorous pollination control at each step in the process. Fresh seed must also be generated every year to supply market demand, and seed delivery systems must be reliable. Due to these constraints and the business opportunity that they represent, most hybrid varieties are delivered by the private sector. Hybrid technology also has an added advantage—it is easily protected as a "trade secret" by simply limiting access to the inbred parents. Due to gene segregation, seeds saved from F1 hybrids have very poor ability to reproduce the traits of interest in the next growing season. This inexpensive form of intellectual property protection is not subject to any legal requirements or restrictions, making hybrid technology valuable and easily controllable. Because development of hybrid seed represents a significant upfront investment, it is often more expensive to purchase and is typically used where farmers have greater access to credit and other kinds of agricultural inputs. However, companies still must maintain competitive pricing in order to attract farmers to purchase their products. Both large and smallholder farmers around the world have widely adopted hybrid varieties, attesting to the success of hybrid breeding.

2.7 Genomics-assisted breeding technology

New DNA sequencing technologies have expanded modern breeding into the field of genomics, the study of entire genomes of organisms. With the availability of high-resolution genomics platforms, there is growing interest in using DNA markers to enhance the efficiency and precision of classical breeding to provide a more comprehensive understanding of a breeding line's potential. DNA sequence differences (polymorphisms) mark particular regions of a chromosome and act as an individual's genetic fingerprint. Marker-assisted selection (MAS) uses DNA polymorphisms to predict performance, often before genes or traits are even expressed (Eathington et al., 2007). Breeders are beginning to use information about thousands or millions of DNA polymorphisms across the genome simultaneously to make genomic predictions about plant performance (Heffner et al., 2009). Scientists also use genomics to search for rare, valuable alleles in large germplasm collections in a process known as "allele mining"(Bhullar et al., 2009). These applications of genomics represent a major competitive advantage in commercial plant improvement for at least three reasons: they increase the efficiency of selection, saving time and lowering the cost of field evaluation; they facilitate the identification of valuable accessions in gene banks; and they provide a competitive advantage in securing IP.

Using high throughput genomics data, breeders can predict which traits a plant carries and infer how well it is likely to perform in the field. This ability represents a major competitive advantage in the context of commercial plant improvement but puts small programs at a disadvantage because of the cost of genotyping large numbers of plants. Expenses are incurred with every genotyping run, and qualified computational and bioinformatics specialists are needed to analyze the data. While large private-sector breeding companies usually do their genotyping in-house,

public-sector breeding operations often contract out to commercial genotyping laboratories that were originally developed to service the health and pharmaceutical industries. Linking with the human health community to optimize access to genotyping services at prices that take advantage of economies of scale, agricultural research programs in developing countries can now evaluate varieties utilizing up-to-date technology platforms adapted for a wide range of plant species. This capacity allows many research institutions to avoid sending plant or DNA samples out of the country, quelling concerns about patenting and ownership of resources. Commercial genotyping laboratories profit by staying abreast of new technologies, maintaining strict quality control standards, keeping costs competitive, and respecting the proprietary nature of data. Ensuring access to reliable, state-of-the-art technical support provides an invaluable service to a global community of publicly funded research institutions—dissemination of technological innovation without creating dependencies on private seed companies.

Genomics technology and MAS have also opened up many new opportunities to utilize rare genes from wild species and landraces located in the world's gene banks, which could improve pest and disease resistance, abiotic stress tolerance, nutritional quality, or yield (Brar and Khush, 2003; Septiningsih et al., 2008; Yan et al., 2010). MAS is also widely used to move transgenes into new varieties (Nishiguchi et al., 1998; Gao et al., 2011).

2.8 Genetic engineering

Genetic engineering involves the use of recombinant DNA technology, rather than sexual crossing, to expand the gene pool available within a given crop species. All transgenic crops on the market today have been developed using a species of bacteria, *Agrobacterium tumefaciens*,

which has the natural ability to insert DNA randomly into the chromosomes of a plant that it infects—a process known as transformation (Klee et al., 1987). It is a common misconception that crop biotechnology involves a simple cut and paste procedure (Gepts, 2002). Because *Agrobacterium* randomly inserts a gene of interest into a plant's genome, transformation events must be screened to confirm that new traits are expressed correctly and no other characteristics have been disrupted. Sometimes multiple transgenes are put into a crop variety, a process known as gene stacking. This can be used to introduce multiple independent traits (Bt and Roundup Ready, for example) or multigenic forms of a single trait (different forms of herbicide tolerance) (Que et al., 2010). Patented traits developed at different times or by different companies can be combined in the same variety, although complex licensing agreements must be established to ensure proper revenue flow.

Crops have varying degrees of transformation efficiency, and often transformation is only possible in a few varieties of a given species (van Wordragen and Dons, 1992). As a result, transgenes are regularly introduced into one variety and later transferred between varieties using sexual crosses and MAS. In asexually propagated crops like potato, sweet potato, and cassava, however, it is difficult to use crosses to improve existing varieties, let alone to move transgenes between varieties. Most yield gains in potato over the past 150 years are not the result of genetic improvement, but have instead come from improved agronomic practices that result in healthier tubers (Douches et al., 1996). Genetic engineering can speed varietal development in asexually propagated crops by bypassing sexual barriers altogether, although transformation protocols must be established for every variety within these species—a hurdle that can be difficult to overcome (Prakash and Varadarajan, 1992).

The precautionary principle and regulation of transgenic crops

Public apprehension about genetically modified organisms (GMOs) is intertwined with moral, religious, political, economic, and scientific concerns that play out in complex ways. Because biotechnology facilitates transfer of traits between totally unrelated species and essentially leapfrogs millions of years of evolution, many liken genetic engineering to "playing God" or tampering with the balance of nature. Others in the anti-GMO movement see biotechnology as an example of wealthy nations, companies, or institutions using patents and trade secrets to block access to technologies, crop varieties, and essential knowledge that are needed in the developing world. Scientific apprehension about genetically modified crops is rooted in awareness of how little we understand the complex ecosystems that sustains life on Earth. Many ecologists and evolutionary biologists believe that formal scientific investigation is warranted to better understand the long-term consequences of what appear to be small changes brought about by genetic engineering. Novel and potentially potent new trait combinations may have emergent properties in complex biological or ecological systems that are not predictable using the reductionist principles of molecular biology (Regal 1996).

All of these concerns have led to the application of the "precautionary principle" and the development of biosafety regulations designed to protect the health of the consumer, the environment, the agricultural system, and the ecosystem as a whole. Each transgenic event must be registered, evaluated, and officially released in the country where the crop is to be grown. The identity and copy number of the individual transgene or genes in the plant genome are carefully documented, and the GMO is evaluated in field trials over years and environments to confirm the identity and stability of the new trait. Multiple government agencies are generally involved in the

regulation of biotech crops, often creating numerous technical and bureaucratic hurdles that impose significant financial costs and delays on the approval process. Regulatory bodies are also required to successfully manage growth, sale, distribution, and trade of GMO seed and produce. Many developing regions of the world have difficulty establishing policies due to lack of trained personnel, political instability, or confusion surrounding the safety of GMOs. Regional efforts are emerging in order to overcome these hurdles, including the East African Community, Common Market for Eastern and Southern Africa, Southern African Development Community, and the *Economic Community of West African States*. It is interesting to note that varieties generated using traditional plant breeding techniques undergo much less stringent testing for food and environmental safety than varieties produced using genetic engineering.

How have GM crops impacted farmers?

GMOs have only been in commercial production in the U.S. since 1996. Since then, global commercial production has increased linearly, from 1.7 million to over 175 million hectares in 2013—representing the fastest adoption of any crop technology in history (ISAAA, 2013). In 2011, biotech crops accounted for roughly 11 percent of the world's agriculturally productive 1.5 billion hectares and are expected to reach 13.3 percent of total production by 2015 (ISAAA, 2011). The U.S. continues to have the largest acreage of GMOs for a single country, 70.1 million hectares in 2013 representing 40% of total production. Additionally, 2013 was the second year in which developing countries surpassed industrialized nations in GMO production. Roughly 46 percent of the world's biotech crops were planted in Brazil, India, China, and Argentina—developing countries that represent about 41 percent of worldwide population (ISAAA, 2013). Biotechnology has had well-documented positive impacts globally: of the estimated 18 million

farmers growing GMOs in 2011, 90 percent reside in developing countries and subsist on less than two hectares of land (ISAAA, 2013).

Nearly all biotech crops in production today are soybean, maize, cotton, and canola varieties engineered for only two traits—insect resistance (Bt) and herbicide tolerance. These traits were designed to reduce manual labor, eliminate the need for highly toxic chemicals, and facilitate conservation agriculture practices. Globally, there has been little incentive to develop GMOs for human consumption because of public opposition to biotechnology. As a result, only a few GM crops in the U.S. have reached consumers as produce in the grocery store; examples include virus-resistant papaya (Tennant et al., 1994), virus-resistant squash (ISAAA, 2011), and Bt sweet corn (Shelton, 2012). While these crops have little impact on global food security, they represent an emerging trend that is likely to pick up in the future.

Bt crystal (cry) toxins are derived from the soil bacterium *Bacillus thuringiensis* and have been used for decades in organic agriculture, due to their safety for both human consumption and the environment. Different cry genes have been engineered into several major crop plants to specifically target certain insects, such as the cotton bollworm. Bt crops control insect pests, reduce pesticide usage, and have dramatically increased on-farm yields; from 1996 to 2008, Bt crops reduced chemical insecticide use in corn by 35.3 percent and in cotton by 21.9 percent (Brookes and Barfoot, 2010). These changes increased income for millions of smallholder farmers. Kathage and Qaim (2012) tracked the economic impact of Bt cotton in India from 2002 to 2008 and showed that smallholder adopters had a 50 percent gain in cotton profit and an 18 percent increase in monetary living standards. However, these changes did not occur without

considerable controversy (Kathage and Qaim, 2012). In the early 2000s, local farmers introduced Bt genes into Indian cotton cultivars from an unknown source of germplasm. The use of Bt cotton became so widespread that it led to the subsequent approval of its cultivation by the Indian government (Herring, 2007). In this case, local farmers were more successful than multinational corporations in battling anti-GMO activists, who were organizing widespread protests, burning Monsanto Bt cotton fields, and spreading unsupported rumors concerning farmer suicides (Herring, 2007; Herring and Rao, 2012). In 2013, over seven million smallholder farmers in India and seven million more in China cultivate Bt cotton. India accounts for one third of global cotton hectarage, 95 percent of which has one or several Bt genes (ISAAA, 2013).

Herbicide tolerant crops have had a similar success story. Glyphosate (Roundup) was discovered in the 1970s and has been heralded as the least toxic and most environmentally benign chemical herbicide. It has a low tendency to bioaccumulate, is relatively immobile in the soil, and is not a major polluter of waterways; it is quickly bound by organic matter and subsequently degraded by microbes, a process that only takes several weeks (Sprankle et al., 1975; Shuette, 1998). Glyphosate inhibits an enzyme that is absent from mammals but that is essential for synthesizing key aromatic amino acids in plants (Amrhein et al., 1980; Steinrücken and Amrhein, 1980); Roundup Ready crops contain a modified version of this enzyme that is resistant to the chemical, so that when fields are sprayed only the crop survives. Glufosinate (Liberty) herbicides contain the active ingredient phosphinothricin, which kills plants by blocking an enzyme responsible for nitrogen metabolism and ammonia detoxification—a by-product of plant metabolism. Liberty-link crops are engineered to express an enzyme that blocks the effects of phosphonothricin. Herbicides have become important in the adoption of CA in mechanized farming systems. In

2012 alone, it is estimated that the use of herbicide-tolerant crops saved 24.6 billion kg of carbon dioxide from entering the atmosphere by facilitating CA practices, equivalent to taking about 10.9 million cars off of the road (ISAAA, 2013).

Biotechnology: A private endeavor

Although genetic engineering is not a magic bullet for the majority of challenges faced by farmers in resource poor regions of the world, it will continue to offer many opportunities for enhancing food security when integrated into sustainable crop production systems. Large companies are willing to invest in biotechnology because patents and use-conditions ensure a constant revenue stream, as utility patents in the US, EU, and Australia force farmers to repurchase seed every year if they want to use a biotech variety. Roundup Ready crops require farmers to purchase herbicide as well as the GM seed, ensuring a double revenue stream. However, farmers are only willing to purchase GM crops and extra inputs as long as the benefits of the GM trait outweigh additional costs incurred. By competitively pricing GM seed, private companies have successfully introduced a technology into the marketplace that is easily identifiable, proprietary, and demonstrably profitable for both industry and farmers.

Development of biotechnology is not a risk-free enterprise. The average cost of researching, developing, and gaining approval to release a single transgenic crop variety is estimated at $135 million (ISAAA, 2011), with no guaranteed payoff in the form of market or production success. This is a significant barrier for both small companies and public sector breeders that want to enter the biotech market, and often only large, well-financed companies have the capacity to take innovative R&D and successfully bring it to the market as a product. Thus, many multinational

seed companies buy out smaller start-up companies that do groundbreaking research and incorporate their R&D into existing biotech development pipelines. This concentration of biotechnology patents within the hands of a few large corporations is no different than the germplasm arms race, and it continues to impede R&D in the public sector from moving forward—particularly when developing crop varieties for humanitarian efforts where no profit is to be gained.

By establishing a presence within the farming community, the private sector has arguably begun to sway public opinion concerning the value in biotech crops—a movement that is visible all over the world. Parts of Europe are now growing biotech maize, and Japan imports transgenic soy and papaya from the US, although it still prevents production of GM crops within the country. As developing regions of the world continue to accept the use of biotech traits as a major means of combatting biotic stressors such as pests and weeds, anti-GM regions that are dependent on importing food will be increasingly likely to encounter transgenics as the major available commodity. Additionally, countries such as China, India, Brazil, and South Korea are generating their own native biotechnology industries and collections of patents. Many of these countries will seek to export technology, while continuing to be major importers of food, fiber, and fuel crops. This new set of players will alter the supply and demand chains all over the world. As observed in cotton, biotech crops may even become the preferred way to meet the global demand for certain agricultural commodities.

2.9 Crop technologies in the coming decade

Next-generation crop varieties targeted to enter the market in the coming decade will be bred for novel traits, including disease and pest resistance, abiotic stress tolerance, and in some cases, nutritional quality. In order to deliver improved varieties in a timely manner, many of these traits will have to be bred using combinations of several technologies, which will require new institutional collaborations, both public and private, on a much greater scale than ever before.

Biotechnology is expected to bring significant improvements to numerous crop species in the coming decade, largely by addressing abiotic and biotic stressors that are currently difficult to manage using traditional means. However, despite increasing adoption rates of biotechnology globally, there are still considerable political and social barriers surrounding the adoption of transgenics that will have to be overcome—especially if transgenics are to reach their maximum potential in addressing issues of food stability.

Genome editing

Recent discoveries in plant pathology offer exciting new opportunities to develop targeted forms of disease resistance that do not depend on the introduction of transgenes. Scientists have discovered that the pathogenic bacteria *Xanthomonas* uses a class of proteins known as transcription activator-like (TAL) effectors to alter gene expression and cause disease in rice, pepper, and other crops (Boch et al., 2009). By deleting DNA sequences targeted by TAL effectors, it is possible to confer immunity to disease (Li et al., 2012). More recently, the CRISPR/Cas9 system has been used to precisely edit the genome of numerous organisms (Mali et al., 2013). CRISPR/Cas9 plasmids can be engineered to target any position in the genome by

using a guide RNA sequence as a template. At the specified location, a DNA digesting domain within the CRISPR/Cas9 plasmid can be used to induce mutations at the resolution of a single base pair (Mali et al., 2013). This exciting development is completely heritable, precisely targeted, and easily bred into subsequent generations of plant varieties. Furthermore, it differs from standard genetic engineering because it carries no footprint of human intervention and is identical to many natural mutations causing small deletions in the genome. A recent study used a combination of CRISPR/Cas9 and TAL-mediated engineering generated wheat resistance to powdery mildew through target mutation of three genes (Wang et al., 2014). Whether these approaches will be subjected to the same biosafety regulations as other forms of genetic engineering has yet to be determined, and public acceptance cannot be predicted. Nonetheless, genome editing is opening new doors for plant breeders attempting to improve traits in crops that are difficult to cross sexually.

Cisgenics

In polyploid, asexually propagated, or perennial crops, such as potato, sweet potato, or apple respectively, genetic engineering will speed up and diversify the deployment of resistance genes. Transforming plants with genes from wild or domesticated relatives is being rebranded as "cisgenics" rather than transgenics (Schubert and Williams, 2006), because genes are coming from sexually compatible species and there is no introduction of "foreign DNA"—though genetic engineering is still involved. Cisgenic transformation is being used to develop apple varieties resistant to apple scab caused by *Venturia inaequalis* (Vanblaere et al., 2011), and potato varieties resistant to late blight caused by the fungus *Phytophthora infestans,* the same pathogen that was responsible for the Irish potato famine in the late 1800s (Boonekamp et al.

2008)(Jacobsen and Schouten, 2008). Stacking multiple disease resistance genes from cultivated or wild relatives in modern varieties using cisgenics will provide durable resistance to diseases without the long and arduous process of conventional approaches to breeding.

Biotech vegetables

Some of the largest gains in yield and nutritional quality in the coming decade will likely come from introducing transgenes into vegetable crops. The amount of arable land devoted to cultivation of vegetable crops is expanding 2.8 percent annually, more than any other type of crop globally. As of 2010, vegetable crops demanded the use of more pesticides than any other major crop—twice what was used in rice, and more than in corn and cotton combined (Shelton, 2012). The relatively high use of pesticides is due to expansion of cultivation and the fact that market success of a vegetable crop is directly dependent on cosmetic appearance. Chemical control of pests and diseases is currently the most common strategy for meeting consumer demands.

Pest resistant Bt potatoes, crucifers, and eggplant have already been developed, but have not yet reached the market due to high costs of regulation, absence of regulatory systems, confusion surrounding the safety of GM crops, and political lobbying. The Genetic Engineering Approval Committee (GEAC) of India approved the cultivation of Bt eggplant in 2010, after rigorous testing showed that it met the "laws of substantial equivalence" and was just as safe to cultivate and consume as non-engineered varieties (Shelton, 2012). Non-engineered eggplant currently experiences 60 to 70 percent losses due to pests and is sprayed an average of 27 times during a growing season. Bt eggplant produces twice the yield of modern hybrids and could provide

economic gains of $108 million annually, due to increased availability of produce to consumers and significantly reduced pesticide usage (Krishna and Qaim, 2008). However, political lobbying by major nonprofits such as Greenpeace led to an indefinite moratorium on Bt eggplant's release. Ironically, the moratorium was invoked citing the precautionary principle, yet has resulted in the continued overuse of deadly pesticides (Kolady et al., 2010).

There are similar stories surrounding the release of Monsanto's NewLeaf potato during the 1990s, which was stacked with Bt genes, a potato Y virus resistance gene, and a resistance gene for potato leaf roll virus. Monsanto ultimately closed its potato breeding program in 2001, after intense political lobbying and boycotts from major U.S food chains like McDonalds. Farmers also contributed to NewLeaf's lack of success because they opposed the requirement to plant refuge areas and purchase new tubers every year (Shelton, 2012). It is noteworthy that the significant public opposition to the use of GM vegetables, even in the US, keeps them from reaching the market—despite the benefits they offer farmers and consumers alike.

Although debates surrounding biotechnology in the coming decade are inevitable, transgenic variety development is continuing at a faster pace than ever—in hopes that the benefits of genetic engineering will ultimately outweigh concerns. Insect-, fungal-, and viral-resistant varieties of cassava, pigeon pea, and sweet potato are being developed for Africa (ISAAA, 2011), and Bt-rice is expected to be released in China in the coming decade (Choudhary and Gaur 2009). While Bangladesh approved Bt eggplant cultivation in 2013, it has yet to be approved in India or the Philippines (ISAAA, 2013).

Biofortification

Nutritional enhancement is a target of both traditional breeding and genetic engineering that requires different strategies for execution, depending on the nutrient of interest. For micronutrients present in the soil, breeding efforts focus on more efficient uptake of nutrients by the roots and partitioning to the plant organs that we ultimately consume (White and Broadley, 2009). Traditional breeding for vitamin-A enriched maize, sorghum, and sweet potatoes, and high iron beans is well-documented. Because many crop varieties do not contain the genes that are responsible for partitioning and accumulating micronutrients and vitamins, biotechnology represents a way of introducing novel nutrient pathways into a crop variety (Naqvi et al., 2009). Iron, zinc, vitamin-A, and vitamin-E enhancement are currently being genetically engineered into cassava, banana, sorghum, rice, and maize (ProVitaMinRice Consortium 2012), with vitamin-A enriched golden rice expected to reach production in the Philippines and other Asian countries by 2013 or 2014 (ISAAA, 2011). While the best long-term solution for addressing micronutrient disorders is clearly diet diversification, delivering critical micronutrients in a food crop for which delivery channels are already well established has support from many sectors.

2.10 Concerns and opportunities for the coming decade

There are many promising new crop technologies in the pipeline that can offer farmers relief from devastating diseases, insects, and environmental stresses, but continued investments in traditional breeding are still a high priority. Although biotechnology has facilitated successful delivery of defensive traits into multiple species, it has yet to develop a variety from scratch or improve yield per se—and thus it is still dependent on traditional breeding. Adapted plant varieties are integrated biological systems whose underlying genetic circuitries are replete with

built-in redundancies, feedback loops, and finely tuned genes. While inserting or modifying a small number of carefully selected genes can alter plant physiological processes and response to plant disease and environmental cues, we need to continue to invest in germplasm enhancement and traditional forms of plant breeding, including genomic-assisted breeding, to maintain a steady flow of improved germplasm that provides the basis for both biotechnology and essential crop production.

As a complement to plant breeding, there is a growing global awareness of the need to manage natural resources more sustainably in agricultural production systems. The major challenges for the coming decade revolve around how new crop technologies will be deployed, who will benefit from them, and what political, economic, and environmental consequences will ensue. National and international policy support will be critical for balancing competing interests as people, institutions, and governments vie for the resources needed to meet the rising demand for food, fuel, and fiber, and cope with the social and ecological disruption associated with climate change and the displacement of human populations. This will require complex, system-level thinking and constant attention to changing realities on the ground.

Genetic uniformity and vulnerability

Wide-scale adoption of hybrids and other modern varieties has been criticized for contributing to the uniformity of crop genetics in farmer's fields (Day-Rubenstein et al., 2005). Genetic uniformity imposes extremely high selection pressure on pests, weeds, and diseases, which quickly evolve to overcome barriers to their survival due to relatively short reproductive cycles and prolific numbers of offspring (Delp, 1980). In some cases, resistance breakdown can occur

within just a few years (McDonald and Linde, 2002a; McDonald and Linde, 2002b; Palloix et al., 2009). A restricted crop repertoire also narrows the nutritional options available to humans, especially when poor infrastructure, geographic barriers, or political pressures restrict trade.

Geographically, crop genetic diversity can be quantified based on the number of varieties in the field at any point in time, the distribution of varieties within an area, and the genetic distance between varieties (number of species) within a given area or period of time (Day-Rubenstein et al., 2005). In traditional farming systems, spatial diversity helped lower the risk of crop loss due to pests, diseases, and variable environmental stress. In modern production systems, temporal diversity represents an alternative risk-mitigating strategy, but requires a pipeline of new, genetically diverse varieties that can be steadily released over time (Duvick 1991). While genetic variation is a key component of ecosystem stability, some forms of genetic variation are more resilient than others. Some crop varieties and some mixtures of varieties are more resistant to pests, diseases, or abiotic stresses than others (Wolfe, 1985; Mundt, 2002). Thus, modern varieties, including transgenics, that are bred to provide multiple and complex forms of pest resistance may provide more durable resistance to pests, pathogens, and weeds than landraces.

The history of modern agriculture is littered with examples of low genetic diversity in crops leading to pandemic loss of yield. The Irish Potato Famine of the 1840s was due to widespread susceptibility to a single pathogen, *P. infestans* (Donnelly 2001). Similarly, the use of a single variety almost led to the destruction of the banana industry in the 1950s due to Panama disease, which is caused by the fungus *Fusarium oxysporum*. A new variety that is resistant to *F. oxysporum,* 'Cavendish', has been in production for over fifty years. However, the industry is

currently under threat again, due to a more virulent strains of *F. oxysporum* a second fungus *Mycosphaerella fijiensis*, and banana bunchy top virus (BBTV); there is currently no variety with resistance to all three pathogens (Dita et al., 2010).

The U.S. Southern corn leaf blight epidemic in 1970 occurred due to the use of a single source of male sterility (T-cytoplasm) in hybrid maize in the US, which caused susceptibility to the fungus *Helminthosporium maydis*—a disease was never before a serious problem. Widespread use of this form of male sterility resulted in a 15 percent loss of the U.S. maize crop in a single season, despite abundant variation in the nuclear genome of maize varieties (Tatum, 1971; Ullstrup, 1972). Today, widespread use of a single source of cytoplasmic male sterility in rice hybrids warrants caution (Cheng et al., 2007). Other extreme examples include: coffee rust, caused by the fungus *Hemileia vastatrix*, which can eliminate 30-90% of yield depending on the year (Bigirimana et al., 2012); a new form of wheat stem rust caused by the fungus *Puccinia graminis*, to which 90% of wheat varieties are currently susceptible (Singh et al., 2011); and maize lethal necrosis disease, which is caused by a combination of two viruses and has resulted in up to 100% crop loss in Kenya (Wangai et al., 2012). Unpredictable and extreme weather events, particularly in temperature and rainfall, often exacerbate the spread of disease, and it is not difficult to imagine scenarios where deployment of a single variety of any crop could result in serious loss with potential to disrupt the entire global food supply.

Similar concerns surround transgenic crops. The Roundup Ready trait has been incorporated into numerous species, and all are dependent on a single chemical herbicide, glyphosate. With the spraying of glyphosate on millions of hectares globally year after year, it is inevitable that

herbicide-tolerant weeds will emerge; indeed, many have already been reported (Powles, 2008). Glufosinate-tolerant crops warrant similar concern. Even Bt-resistant crops are dependent on just a handful of cry genes conferring insect resistance. Insect-resistance to Bt has also been reported (Tabashnik et al., 2009), although the rate of resistance has been far slower than with most conventional insecticides (Bates et al., 2005), thanks in part to the mandated use of refuges.

Using genetically diverse germplasm in breeding can greatly reduce uniformity in modern crop varieties. Genomics can facilitate this process by allowing breeders to selectively introgress traits from diverse sources without disrupting the adapted gene complexes and quality traits that have been painstakingly developed over many decades. Placing value on enhancing genetic diversity within a breeding program and deploying diversity at the landscape level introduces a long-term perspective that is key to mitigating the effects of climate change and coping with new pest and pathogen populations.

At the landscape level, the release of new hybrid varieties can be orchestrated to maximize genetic variation over both space and time. Because farmers must purchase F1 hybrid seed every year or every other year, if competitive varieties carrying diverse forms of resistance to insects, weeds, and diseases are developed using numerous hybrid genetic backgrounds, they can be strategically released into the market to ensure that farmers plant a mosaic of different hybrid varieties every year. Diversity in the marketplace is the surest way to avoid uniformity in the field.

Conservation agriculture also contributes to crop diversity by emphasizing the use of crop rotation, which ensures different species are planted in succession over the years in the same field. Although seed companies and smallholder farmers do not reap any short-term benefits from maximizing diversity, and thus cannot be expected to prioritize this when deciding what to put in the field, it may be possible for policymakers to introduce incentives supporting crop diversity. Policies and interventions that encourage the use of spatial and temporal variation in cropping systems should be investigated as a critical and cost-effective policy contribution to food security.

To slow development of super weeds and super pests, modern transgenics has stacked transgenes. In 2013, roughly 27% of hectarage dedicated to biotech crops contained stacked transgenes (ISAAA, 2013). Gene stacking is often combined with planting non-transgenic, pest-susceptible refuge areas, practices that are required by law in most countries growing transgenics. It is important to maintain refuge areas, even in the presence of gene stacking, as non-GM areas help to lower the pressure on populations of susceptible insects and weeds to develop resistance(Bates et al., 2005). Durable pest- and disease-resistant varieties are also in the pipeline.

The technology gap and agricultural R&D

Most of the crop technologies discussed in this chapter have been around for decades and could be easily applied to numerous crop species and scenarios with obvious advantages for smallholder farmers in the developing world. Why is this not happening? What is needed to

catalyze more effective integration of crop technologies with the farming practices of smallholder farmers?

More than 90 percent of global agricultural research is conducted in developed countries and is focused on the crops important in those economies. The private sector accounts for more than half of those R&D expenditures, influencing the kind of research and IP protection that emerges from investments (Pardey et al., 2006). For every $100 of agricultural output, developed countries invest eight to ten times the money on research and development—and they do so with an expectation of return on that investment. In developing countries, public funds are still the major source of support, and the private sector accounts for just six percent of R&D (Pardey and Pingali, 2010). These figures underscore the different phases and philosophies of agricultural R&D that separate developing and developed countries. There is enormous potential for agronomic and genetic improvement and rapid market development for many orphaned indigenous food crops, such as those that feed a large percentage of the African population: yams, pigeon pea, millet, sorghum, and cassava. Currently, the private sector is the major source of multinational investment in agricultural R&D, but because these crops are deemed to hold little commercial promise, they remain a low priority for investment.

Yet industry is motivated to invest when there are markets that offer a reasonable return on investment. Technology alone cannot motivate development; a supportive environment is a prerequisite. In most developing countries today, the level of public investment in agricultural R&D is too low to foster and maintain a skilled workforce, making private investment unlikely to come to the rescue. There has to be enough public investment in place to support local

infrastructure, as well as skilled people who can adapt new technologies to local needs. If GM crops are involved, the existence of appropriate biosafety legislation and administration is essential. Potential market size and an economic and intellectual infrastructure that can support and motivate a high level of innovation and entrepreneurship are also crucial.

Beyond patent protection

Historically, open sharing of germplasm and technology has been the first step to developing a vibrant agricultural R&D sector (Smolders, 2005). The explosion of patenting in the developed world, the pace of discoveries, and the high level of investment in the biological sciences that underpin new agricultural technologies have created barriers to innovation that continually marginalize those most in need. New forms of public-private partnerships that protect assets accumulated with public funding are urgently needed, along with reconsideration of the utility patent system as it relates to plant variety development.

Adopting Plant Variety Protection as a form of IPR in most developing countries is more appropriate than adopting a utility patent system for crop varieties, at least during the early phases of agricultural development. Ensuring breeders' exemption and farmers' privilege can facilitate access to germplasm and reward local breeders for sharing and utilizing efficient breeding technologies, rather than seeking to protect or own them. Allowing utility patents to enter the scene too early is likely to impede crop technology development that is essential to addressing smallholder needs (Smolders, 2005). The history of corn breeding in the U.S. suggests that the absence of strong IPR played a critical role in facilitating the development of the hybrid corn industry. Initially, all hybrid corn companies had access to the same public

inbreds, and each company self-pollinated superior hybrids of their competitors to develop new inbreds (Troyer, 1999; Troyer, 2004). Thus, hybrid varieties and inbred lines are essentially derived from varieties developed by public sector breeders in multiple countries. Gouache (2004) argued that if the intellectual property practices of today had been in place 50 years ago, it is very unlikely that U.S. corn yields would have reached today's level.

Indeed, crop improvement was and largely remains a cumulative process. New varieties are built upon past selection, and the progressive nature of this process means that past discoveries and related research are an integral part of contemporary innovations. Current intellectual property laws created to protect the linear process of developing a single crop technology do not reflect this cyclical need to return to "old" germplasm resources or technologies when developing new varieties.

Although patents ultimately exist to motivate innovation and disclosure, too many competing patent rights surrounding an invention can lead to less innovation and may actually prevent useful and affordable products from reaching the marketplace. This situation, referred to as the "tragedy of the anti-commons" (Heller and Eisenberg, 1998), has become pervasive in the world of biotech crops. It is often challenging to cut through the dense web of overlapping IP rights surrounding an invention. Given the pace of technology development in genetic research, there are often dozens or hundreds of patents held by multiple companies relating to a single variety. These "patent thickets" require users and other innovators to make licensing deals with multiple patent holders to commercialize new products that build on prior inventions. The transactions add significant costs to the licensing process that are passed on to consumers and frequently

prevent entry into new markets. These unintended consequences of the current patent system stifle competition and are especially insidious when they prevent the benefits of technology from trickling down to the public breeders that address the needs of smallholder farmers.

Collaborative and open-source licensing

In cases where utility patents are already in place or cannot be avoided, as is often the case for biotech crops, cooperative licensing can help newcomers navigate through the patent thicket, lower transaction costs, hasten innovation, and commercialize new technologies (Shapiro 2001). Patent pools, clearinghouses, and open-source licensing are examples of cooperative licensing strategies that have been used in the international plant breeding arena to facilitate access to new genetic technologies for humanitarian purposes.

A patent pool is an agreement between two or more patent owners to license one or more of their patents as a package to one another or third parties that are willing to pay the associated royalties (Merges 2001). Patent pools could clear the patent thickets and lower the barrier to entry for new technology users. An example is the Golden Rice pool, where six key patent holders agreed to grant licenses for more than 70 patents, free of charge, to developing countries. This pool was established by a public-private partnership that formed a single licensing authority to speed the development of biofortified golden rice varieties for dissemination to farmers in Africa and Asia. Market segmentation may also be negotiated as part of a patent pool. The poorest smallholder farmers can obtain improved biofortified rice varieties at minimum cost, while those who can afford to pay shoulder higher costs.

A clearinghouse is a mechanism whereby providers and users of goods, services, information, or patents are matched with each other to improve access to patented technologies (Hope 2008, 2009). An example of a clearinghouse is the Public Intellectual Property Resource for Agriculture (PIPRA). PIPRA's goal is to mobilize patented technologies from a wide range of public or private technology providers to address specific projects for the improvement of subsistence and specialty crops. They seek to make agricultural biotechnology inventions more accessible to emerging countries by providing one-stop shopping for IP information, supporting the development of IP management best practices, enhancing IP management capacity in developing countries, facilitating access to public sector patented technologies, and developing gene transfer and gene-based trait technologies that have maximum legal freedom to operate.

Open-source licensing means anyone, anywhere, is allowed to copy, modify, and distribute the technology for any purpose, without having to pay royalties to the owner of the license, with the stipulation that any improvements made to the technology must be shared under the same open-source license. A relevant example is the Biological Innovation for Open Society (BiOS) License from the Centre for Applications of Molecular Biology in International Agriculture (CAMBIA) (www.bios.net/daisy/bios/mta/license-intro.html). The CAMBIA-BiOS initiative has developed a public-access database of patents in the life sciences to help users analyze the IP landscape, navigate IP thickets, acquire freedom to operate, and forecast trends and new technology; created new genetic technologies and delivered them through open-access licenses; and developed new mechanisms for licensing and contract agreements to influence national and international policy and encourage democratized problem solving.

Open source initiatives occupy an important niche in the complex IP landscape today. Open-source business models pose different advantages, opportunities, and trade-offs than do proprietary models; co-existence of both can drive competition and promote efficiency and innovation in the marketplace (Lerner and Schankerman, 2010). While the two approaches are historically perceived as competitors, the same organization often engages in both forms of IP simultaneously. This suggests that a dynamic combination of the two strategies could be highly synergistic in developing and delivering crop technologies to small-scale farmers in the developing world. Acknowledging and valuing this combined approach opens up new opportunities for public-private partnerships.

Open-source breeding

One example would be an open source collaborative breeding program, whereby a central research body, such as a CGIAR center, collaborates with national program partners and/or small seed and breeding companies. In an era where hundreds or thousands of varieties can be genotyped more quickly and economically than they can be phenotyped in the field, CGIAR centers can provide genotyping services while taking advantage of partners' investments in phenotyping across years and environments. In exchange for information about phenotypic performance, the CGIAR center integrates genotypic and phenotypic information and provides predictions about the performance of new lines and decision support for the selection of future parents and crosses. The collective information is much greater than the sum of the parts: the genomic information can be leveraged to provide more accurate predictions about performance than any single breeding program could hope to produce. Open-source information sharing

would be compatible with other forms of IP protection allowable on improved varieties emerging from the program.

The breeding efforts of the CGIAR center would focus on validating the predictions, developing new breeding lines, and improving source populations to address deficiencies identified broadly across its client breeding programs. Open source breeding could be particularly valuable when breeding for stress-prone environments. Lines evaluated in specific mega-environments share information on a target environment with other members of a cohort (breeding population). Through data pooling, a number of small but cooperating breeding programs could leverage each other's efforts to increase breeding efficiency. IP on successful varieties derived from the program would include a negotiated royalty flow back to the CGIAR to support its activities.

Legislative support of orphan crop R&D

The ultimate goal of cooperative licensing schemes is to circumvent IPRs surrounding crop technologies and facilitate application of successful methodologies. However, instead of relying on complex institutional arrangements to bypass IPRs, policy incentives could be designed to encourage the private sector to invest in orphan crop species that are important for food security, but that have yet to become economically viable. For example, the 1983 US Orphan Drug Act (ODA) was established in order to provide incentives for the private sector to develop treatments for rare medical conditions (defined as <200,000 afflictions annually). Policy makers in the medical community recognized that research and clinical trials for orphan diseases often cost more than the small revenue streams these drug development pipelines are expected to deliver, and that without intervention many orphan diseases would be left out of R&D pipelines. The

ODA outlined clinical research grants for orphan diseases, set up FDA counselling concerning drug development, established a 50% tax credit on clinical trials for orphan drugs, and guaranteed seven-year market exclusivity, regardless of whether or not the drug became a blockbuster success (Szenberg and Lall, 2012).

Before implementation of the ODA, the rate of orphan drug development was the same as non-orphan drugs. By 1998, there were five times as many orphan drugs on the market and only twice as many non-orphan drugs (Lichtenberg and Waldfogel, 2003), Several drugs developed under the ODA, including Cialis, Abilify, and Botox, have had incredible commercial success (Wellman-Labadie and Zhou, 2010). From 1983 to 2010, the FDA approved 353 orphan drugs and granted orphan status to 2,116 compounds. However, this staggering level of innovation has not come without trade offs. Out of 7,000 orphan diseases, only 200 had become treatable by 2010, and many criticize the ODA for encouraging the private sector to develop expensive drugs for exclusive markets (Armstrong, 2010).

An "Orphan Crop Act" could stimulate private investment in neglected areas of agricultural development, while prioritizing the needs of farmers in the developing world using policy incentives. Research tax credits, rights to exclusivity in domestic or foreign burgeoning markets, and subsidies that reduce the cost of field trials could all be used as bargaining chips. To ensure that technologies are actually reaching smallholder farmers, subsidies or tax breaks could be based off of the number of farmers using new varieties, the acreage over which orphan crops have been deployed, or the level of diversity available in local markets (i.e. the number of different crop species or varieties deployed by a company within a given year or area).

Alternatively, in cases where transgenes have already been approved in non-orphan crops (i.e. Bt genes), patents surrounding these technologies could be extended in exchange for deployment of these transgenes into orphan crop species. Providing incentives for the private sector to invest in the developing world would usurp the need to establish complex PPPs every time patents are involved in the development of a new variety.

Capacity building and institutional change

Several new players, including Brazil, China, India, and South Korea, have developed productive, innovative, self-sustaining agricultural research programs over the last 30 years and are poised to provide new contributions to agricultural R&D. International activities funded by foundations and other donors are facilitating access to high throughput genotyping, phenotyping, and genetic engineering technologies on a wide range of crop species (Generation Challenge Program 2012), but technical advances have outpaced the availability of adequately trained personnel in all areas of plant breeding, creating a large gap in the ability to integrate and use these technologies. Most of the expertise needed is currently found in the private sector, where cost-effective integration of new and older breeding technologies is farthest advanced. Universities are having a hard time keeping up with the demand for new breeders, and most qualified graduates are snatched up immediately by the private sector. New training initiatives are urgently needed to prepare breeders, managers, and extension personnel for the complex world they will inhabit. Leveraging the comparative advantage of the private sector to help train people is worth exploring. However, technical expertise is not enough; these professionals must also be grounded in the realities of the developing world and have a passion and commitment to

translate technical potential into crop technologies that will contribute to the well-being of smallholder farmers in the developing world.

Any decision to integrate modern crop technologies into breeding programs within developing countries will take time and require the evolution of an entirely new institutional culture. Genomics can help breeders more efficiently utilize the raw genetic resources that abound in many developing countries. To be successful, however, open-source breeding programs will require reorganization of management structures and reporting protocols (Eathington et al., 2007). The technology itself imposes transparency and rigor, requiring goals and objectives to be explicitly spelled out, time frames set for each activity, both progress and impediments tracked and monitored, and information kept transparent at all levels.

Even as government spending on agricultural R&D is outstripped in many countries by investments in health, education, and other social services, rational policy decisions by national governments could encourage the emergence of new regional centers or support national research centers that already exist in order to create economies of scale and attract multinational agricultural R&D investment (Pardey and Pingali, 2010). Regional centers should be designed to increase access to and dissemination of technical expertise, harmonize the regulatory environment for release of new crop varieties and management practices, and prevent redundancy in agricultural research among neighboring countries. They could operate on a smaller scale than CGIAR centers and coordinate more targeted research goals including: local agricultural extension; maximization of genetic diversity across landscapes; and multi-location field trials with institutions that share similar environmental, abiotic, or biotic stressors. From a

political perspective, regional centers would attract larger R&D investments and could help address food security problems before they become politically destabilizing.

2.11 Toward productivity and sustainability

Accelerating development and deployment of crop technologies could close the agricultural productivity and sustainability gap for millions of small landholders in Asia and Africa and represents a promising approach for avoiding major disruptions in the global food system. In order to meet this challenge, new kinds of public-private partnerships and international collaborations are needed to bring together complementary skills and resources from the public sector, the private sector, government organizations, NGOs, international organizations, and foundations. The emerging economies of Brazil, China, India, and South Korea represent new sources of crop technologies and new engines of growth for international development. These nations are potential investors and should have a stake in global stability, but will compete for natural resources to address concerns about national food security.

The world's international germplasm repositories are essential to plant breeding in the U.S. and around the world and are critical to sustaining the world's food supply. They are irreplaceable and must be supported. Communities of skilled local breeders, agronomists, extensionists, and entrepreneurs that can contribute to closing the yield and sustainability gap in their own countries are essential to economic growth and political stability. In many cases, encouraging the use of Plant Variety Protection rather than utility patents is warranted to promote sharing of breeding technologies and germplasm and to drive innovation. The creation of regional centers for training and research in agricultural R&D can help attract new people and expand the talent pool,

harmonize biosafety standards, facilitate the introduction of new crop technologies, and develop new kinds of partnerships and licensing options that will open the door for greater international investment and economic growth.

CHAPTER 3

CONSIDERATIONS WHEN DEVELOPING AND DEPLOYING A PHENOTYPING TECHNOLOGY

3.1 Introduction

As discussed in the previous chapter, the definition of a crop technology is largely contextual to the target audience in question. For farmers, a crop technology may be as simple as a mechanical plow or a new variety with resistance to a pest or pathogen. However, for plant breeders attempting to develop and release new varieties, the phrase "crop technology" is usually used to describe technological or methodological innovation that advances understanding of the plant species being investigated. Historically, the greatest limitation in modern plant breeding has been access to and characterization of varied germplasm.

When characterizing plant performance, breeders usually view plant traits (phenotypes) as the outcome of an interaction between two major factors: genetics and environmental variation (GxE). Over the past two decades, advancements in molecular biology and DNA sequencing technologies have greatly increased the ability to characterize and understand fundamental aspects of biology across many organisms. In many ways, we are in an era of biological renaissance. While characterizing DNA sequence or RNA expression profiles using next-generation sequencing is technologically sophisticated and requires significant computing and statistical power, the cost of these experiments has dropped dramatically, even in the time it took to complete this dissertation. While genome assembly is still difficult in polyploid species like wheat and potato, for a diploid species like rice it is no longer an issue. *Oryza sativa* has a high

quality, full genome sequences constructed using bacterial artificial chromosomes (BACs) (International Rice Genome Sequencing Project, 2005), and over 3,000 genomes assembled using next-generation sequencing technology (3,000 rice genomes project, 2014). Additionally, public access to real-time weather tracking through satellite imaging, cellular telephones, and ground level remote sensing is also increasing. While these technologies are newer and more expensive to access, it is now possible to accurately measure environmental variation at a range of resolutions: globally, regionally, and in a single field plot. With the ability to collect high-resolution data on both G and E, the final barriers for breeders is, ironically, accurately measuring plant phenotypes.

Plant phenotypes are diverse throughout development, unique to genetic backgrounds within and across plant species, and variable in response to different environments at both the local and landscape level. As such, phenotyping technologies are usually targeted by scientists or breeders to answer specific questions, and often vary in both implementation and scale of data collection (Clark et al., 2012; Tanabata et al., 2012; Cobb et al., 2013). Physiological phenotypes are increasingly being collected at the plot level using handheld, plow, or aerial imaging that quantifies overall vegetative health in the field. Infrared and multispectral cameras can detect plant tissue fluorescence and radiation reflectance signatures as a means of estimating canopy density, photosynthetic efficiency, nitrogen use, and presence of disease (Araus and Cairns, 2014). When plot level plant performance data is paired with satellite geographic information systems (GIS) data on land slope, soil type, and watershed areas, precision agricultural management techniques can be used to deliver targeted fertilizer or water regimens that enhance yield performance and minimize environmental impacts (Mulla, 2013). Automated drone

technology also shows promise in improving pesticide and herbicide management, harvest efficiency, and diagnostic phenotyping of both plants and soil (Bloss, 2014). As these technologies continue to converge, plant breeders will be empowered to develop and deploy plant varieties tailored to specific environmental niches within a field. Understanding how small, phenotypic changes in plant anatomy translate into improvements in plot level performance will be crucial when targeting selection pipelines.

Plant morphological traits, such as branching in rice panicles, are intimately linked with physiological performance, yet are difficult to quantify using plot level imaging because of overlapping structures within an image frame. Photographing individual plants simplifies imaging protocols by reducing background noise, forcing structures into a two-dimensional plain. This has the added benefit of freezing biological specimens in time, and future image-processing techniques can leverage information within photograph databases in much the same way that modern botanists reference dried samples in a herbarium. Additionally, because digital photography is cheap and portable, plant breeders can phenotype in one location and perform analyses in another; this is crucial when working on breeding trials in different locations. Finally, pixel to length conversions are much more accurate than manual measurements, reducing error and increasing the number of traits that can be measured simultaneously. For these reasons, we viewed image analysis as an excellent solution for measuring panicle architecture in rice and set out to develop a novel phenotyping technology, PANorama (Crowell et al., 2014).

3.2 Software development

While image-based phenotyping allows simultaneous collection of numerous measurements, not all measurements are necessarily meaningful in the context of biology. Sophisticated thresholding and skeletonization techniques have made it possible to accurately measure plant organs in photographs (Falcão et al., 2004; Clark et al., 2012; Tanabata et al., 2012; Crowell et al., 2014), but converting skeletons into meaningful phenotypes requires a biological lens. Additionally, although image analysis is faster than manual phenotyping, staging photographs represents a significant time investment. As such, choosing what phenotypes to measure is often just as important as accuracy in methodology, because the tradeoff between the number of phenotypes and the number of replicates evaluated is usually the rate-limiting factor for plant breeders. Chapter 4 describes PANorama's phenotyping capabilities, including a detailed explanation of why the skeletonization algorithms are more accurate and faster than other platforms that are currently available (Crowell et al., 2014). However, PANorama's phenotyping flexibility stems from a careful design of the pipeline as a whole; the skeletonization techniques used to process images are only a piece of the story.

One of the key features we built into PANorama was the ability to customize measurements. Users define phenotypes within PANorama by clicking on an image with a graphical user interface (GUI). Thus, points can be flexibly placed along different intervals within an image to measure different plant structures (Crowell et al., 2014)(see Chapter 4, Methods). The decision to use three points within the interface was strategic; choosing three points is still relatively quick when processing images, yet allows division of an axis into three major organs (root, shoot, inflorescence). Other small features were also designed to facilitate ease of use in the

pipeline. When clicking between images, users do not have to leave the GUI, and the image names are displayed within the window to make it easy to track which samples are being processed. Phenotype thresholds are customizable and easy to edit within an accessory file. Photography in PANorama is controlled by open-source software, gPhoto2, making image acquisition free and accessible to laboratories using different camera setups. Additionally, the linear structure of the PANorama pipeline is more flexible than a self-contained program, allowing users to repeat individual processing steps without having to redo entire analyses. All of these features contribute directly to the flexibility of PANorama, and were the result of beta-testing on thousands of images before public release of the software (Crowell et al., 2014).

3.3 Developing a public phenotyping resource

Because uptake and success of a crop technology is often dependent on dissemination, I wanted to ensure that PANorama was widely accessible to different types of users so that it could become a resource with long-term applicability for a wide range of phenotyping needs. In the short term, that meant releasing an open-source publication a providing the source-code freely so that other researchers can expand on PANorama's functionality (Crowell et al., 2014). In order to do so, I registered PANorama on a public website that hosts open-source software (Table 3.1), which allows tracking of download statistics and can host future releases of PANorama once Chapter 5 of this dissertation has been published.

However, when releasing PANorama's source code at this webpage, I recognized that many of users who are interested in phenotyping software are actually biologists or plant breeders—not computer savvy programmers. To maximize ease of use, I developed in-depth installation

Table 3.1. Summary of webpages referencing PANorama1.0.

Websites hosting PANorama downloads	http://sourceforge.net/projects/panorama1/
	https://vimeo.com/cornellricelab
Websites linking to PANorama	http://ricediversity.org
	http://ricelab.plbr.cornell.edu/
	http://www.plant-image-analysis.org/

instructions targeted for biologists that are included on the webpage's "Wiki" tab. This required integrating a description of the software's capabilities with basic tutorials on installation and executing scripts via the command line. The PANorama1.0 download contains demonstration images for users to test the pipeline, and the webpage depicts a graphical representation from Crowell et al., 2014, which functions as a quick summary of PANorama's core functions (Figure 3.1). To facilitate further uptake of the software, I developed a series of three videos demonstrating usage of the PANorama pipeline (Table 3.1). In order for these how-to videos to be informative, I designed them to contain concise and detailed descriptions; the total viewing time of all three videos is less than 25 minutes, and the files are small enough that they can be easily downloaded and viewed at a later date offline. Finally, I linked these resources across various webpages to make PANorama discoverable (Table 3.1).

3.4 Concluding remarks

When we set out to design PANorama, I didn't want to create a phenotyping technology that was only applicable for our own needs and that would quickly become out of date. The skeletonization algorithms implemented within PANorama were published over ten years ago (Falcao et al., 2002; Falcão et al., 2004), and the fact that these techniques had never been implemented within the plant sciences was, to me, a textbook example of a technology gap preventing progress. Thus, the real innovation behind the development of PANorama was the combination of approaches we used: 1) improve upon a technology that already exists, 2) target and test it for a specific purpose, and 3) package and deploy the platform *with resources* to facilitate its uptake. In practice, researchers often skimp on the third step, which prevents innovation from empowering change. In the 5 months since its release, PANorama1.0 has been

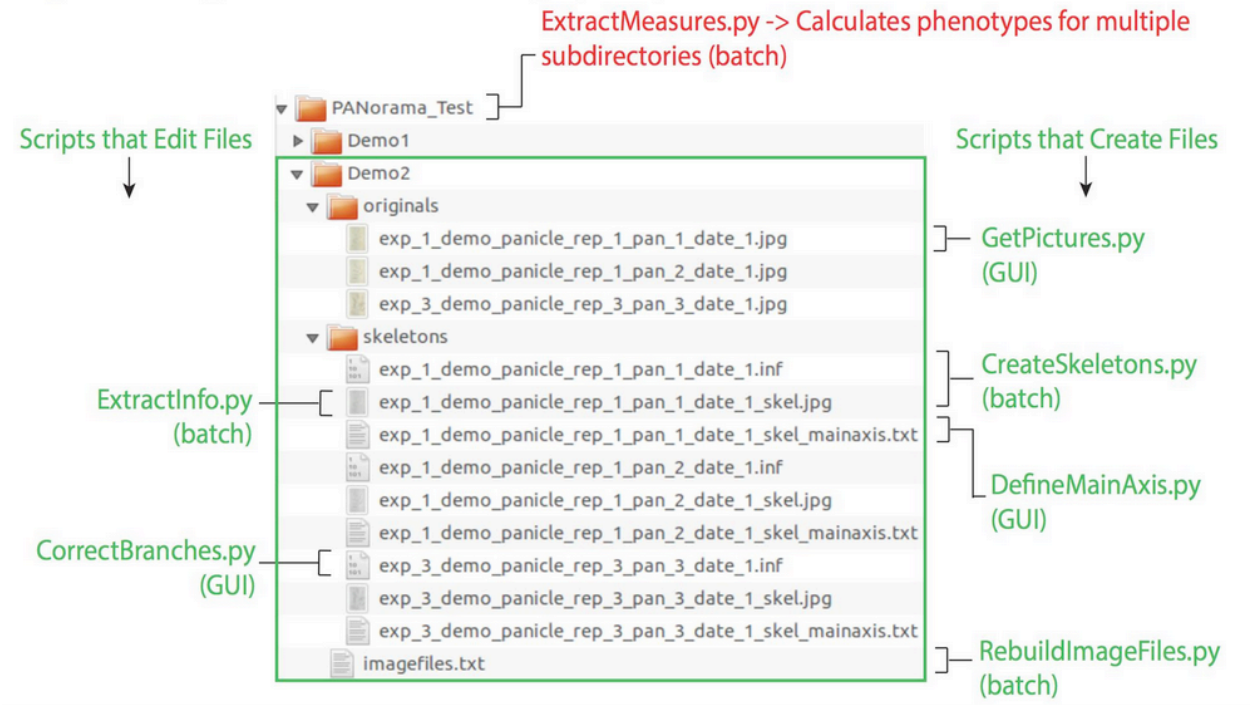
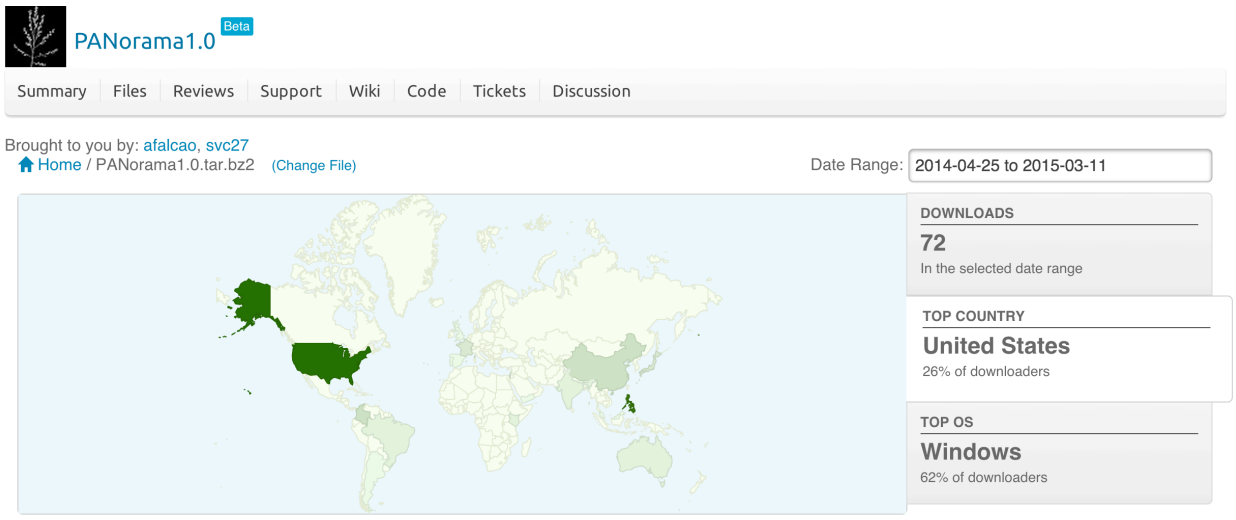


Figure 3.1. Directory structure used in the PANorama Pipeline. This example depicts the master directory /PANorama_Test, with two subdirectories: /PANora- ma_Test/Demo1 and /PANorama_Test/Demo2. All scripts labeled in **GREEN** are executed at the subdirectory level--in this example at /PANorama_Test/Demo2. Scripts depicted on the **RIGHT** side of the folders create files, while scripts on the **LEFT** side of the folders edit files. Scripts labeled **(GUI)** create or edit individual files one at a time, while scripts labeled **(batch)** operate across multiple images. GetPictures.py captures an image and stores it as a .jpg file within the sub-subdirectory "originals." For every image within "originals," CreateSkeletons.py batch generates two files (a .inf and a _skel.jpg file) for every "originals" image and stores them in the sub-subdi- rectory "skeletons." After selecting and saving the three developmental points for one image using the DefineMainAxis.py GUI, a file with the position of the points is generated (._mainaxis.txt). ExtractInfo.py batch edits all _skel.jpg edits to generate final skeletons, using the _mainaxis.txt file of each image. If a user corrects a skele- ton using CorrectBranches.py, only the .inf for that skeleton is edited. If a user executes RebuildImageFiles.py, a new imagefiles.txt file is generated listing the names of all images stored within "originals." Once all images within a subdirectory have been processed, the user can extract phenotypes. The **RED** script ExtractMeasures.py is executed one directory higher, at /PANorama_Test. ExtractMeasures.py calculates phenotypes for all _skel.jpg files stored in both /PANorama_Test/Demo1 and /PANora- ma_Test/Demo2. Phenotypes are stored as .csv files within /PANorama_Test (not shown).

downloaded 72 times in 20 countries (Figure 3.2). The how-to video series has over 209 combined views, and I have personally communicated with researchers using PANorama1.0 at the International Rice Research Institute (IRRI) and at several public universities in the United States. While interest in PANorama stems from the fact that it is a flexible and powerful program, the broad dissemination and uptake of our methodology is clearly the result of resource development and accessibility. This is undoubtedly true for the success of any crop technology.



	Country	Linux	Macintosh	Unknown	Windows	Total
1.	United States	11%	26%	16%	47%	19
2.	Philippines	11%	0%	0%	89%	18
3.	China	0%	0%	0%	100%	4
4.	Colombia	25%	25%	0%	50%	4
5.	Israel	100%	0%	0%	0%	3
6.	France	100%	0%	0%	0%	3
7.	Japan	0%	0%	0%	100%	3
8.	Brazil	50%	0%	0%	50%	2
9.	Korea	0%	0%	0%	100%	2
10.	India	0%	0%	0%	100%	2
11.	Australia	50%	0%	0%	50%	2
12.	Kenya	100%	0%	0%	0%	2
13.	United Kingdom	0%	0%	0%	100%	1
14.	Europe (specific country unknown)	0%	0%	0%	100%	1
15.	Portugal	0%	100%	0%	0%	1
16.	Viet Nam	0%	100%	0%	0%	1
17.	Malaysia	0%	0%	0%	100%	1
18.	Argentina	100%	0%	0%	0%	1
19.	Yemen	0%	0%	0%	100%	1
20.	Spain	0%	0%	0%	100%	1
						72

Figure 3.2. Screenshot from PANorama1.0 webpage (http://sourceforge.net/projects/panorama1).

CHAPTER 4

HIGH-RESOLUTION INFLORESCENCE PHENOTYPING USING A NOVEL IMAGE ANALYSIS PIPELINE, PANORAMA

Included with permission from:

Samuel Crowell, Alexandre X. Falcão, Ankur Shah, Zachary Wilson, Anthony J. Greenberg, and Susan R. McCouch. (2014). High-Resolution Inflorescence Phenotyping Using a Novel Image Analysis Pipeline, PANorama. Plant Physiology, 165(2): 479-495.

Abstract. Variation in inflorescence development is an important target of selection for numerous crop species, including many members of the Poaceae (grasses). In Asian rice (*Oryza sativa*), inflorescence (panicle) architecture is correlated with yield and grain quality traits. However, many rice breeders continue to use composite phenotypes in selection pipelines, because measuring complex, branched panicles requires a significant investment of resources. We developed an open-source phenotyping platform, PANorama, which measures multiple architectural and branching phenotypes from images simultaneously. PANorama automatically extracts skeletons from images, allows users to subdivide axes into individual internodes, and thresholds away structures, such as awns, that normally interfere with accurate panicle phenotyping. PANorama represents an improvement in both efficiency and accuracy over existing panicle imaging platforms, and flexible implementation makes PANorama capable of measuring a range of organs from other plant species. Using high-resolution phenotypes, a mapping population of recombinant inbred lines, and a dense SNP dataset, we identify the largest number of quantitative trait loci (QTL) for panicle traits ever reported in a single study. Several areas of the genome show pleiotropic clusters of panicle QTL, including a region near the rice Green Revolution gene *SD1*. We also confirm that multiple panicle phenotypes are distinctly different among a small collection of

diverse rice varieties. Taken together, these results suggest that clusters of small effect QTL may be responsible for varietal or subpopulation-specific panicle traits, representing a significant opportunity for rice breeders selecting for yield performance across different genetic backgrounds.

4.1 Introduction

Flowering is arguably the pinnacle of plant development. For plants, floral morphology and complex inflorescence architectures are evolutionary embellishments on an age-old process—reproduction and the transfer of genes. For humans, flowers fulfill diverse roles in both food crops and ornamental plant production. This is especially true for cereals (Poaceae), which have remarkably diverse inflorescence architectures that have been the target of selection for 10,000 years (Doust, 2007; Kellogg, 2007).

As the first crop species to be sequenced and a major source of grain for human consumption, domesticated Asian rice (*Oryza sativa*) represents an excellent system in which to study inflorescence development. The rice inflorescence is classified as a panicle, a conically shaped compound raceme in which spikelets are born on long lateral branches instead of directly on the rachis axis (Ikeda et al., 2004). The beginning of rice inflorescence development, known as panicle initiation (PI), occurs when the shoot apical meristem (SAM) transitions into an inflorescence meristem (IM). Over ~7 days, the IM undergoes several transitions into primary branch meristems (PBMs) before ultimately aborting (McSteen, 2006; Yoshida and Nagato, 2011). Rice PBMs are capable of dividing into secondary or tertiary BMs (SBMs, TBMs) before transition into spikelet and floret meristems, resulting in iterative branching morphologies. After IM abortion, floret development and elongation of axes occurs over

roughly 30 days until heading—the point at which the panicle emerges to begin reproduction (International Plant Genetic Resources Institute and Association, 2007).

Every aspect of panicle development in rice is pleiotropically interconnected, coordinated by a host of transcription factors and hormones that are responsive to environmental fluctuation. Variation in plant spacing, day length, temperature, nutrient availability, and water supply can significantly affect panicle development, stunt branch formation, reduce spikelet number, and/or result in sterility—long before a farmer or breeder even sees the first panicle at heading (Kobayashi et al., 2003; Li et al., 2003; Latif et al., 2005; Liu et al., 2008; Thakur et al., 2010). Because anthesis and grain filling occur in a wave of development, beginning with spikelets located furthest away from the rachis axis at the tips of panicle branches and working inward, the position of a spikelet within a panicle determines the time of both fertilization and starch deposition during grain filling. Thus, panicle size, shape, and branching patterns are thought to directly impact grain uniformity and quality (Yoshida and Nagato, 2011). Additionally, tradeoffs between sink and source tissues in rice often translate into direct compensation in the number of seeds or grains per plant (Ohsumi et al., 2011). For example, although large panicles often carry more spikelets, overall seed or grain size usually diminishes. Panicle length and panicle number are also usually negatively correlated. All of these developmental characteristics are directly dependent on the genetic background of the variety and the environment in which it is grown. Currently, the best way to ensure optimal panicle development in rice, and thus stable yield performance, is to strictly control irrigation and fertilizer regimens from several weeks before PI begins until grain filling initiates (Chen et al., 2007). This is true even in modern varieties.

The inherent genetic and morphological diversity for many traits within rice is partitioned into distinct subpopulations. *Oryza sativa* is comprised of two subspecies, *Indica* and *Japonica*, which can be further divided into 5 subpopulations (*indica, aus*, *tropical japonica*, *temperate japonica*, and *aromatic*) (Garris et al., 2005; Zhao et al., 2011). In many cases, novel polymorphisms (and sometimes entirely novel genes) are subpopulation specific (Xu et al., 2006; Fukao et al., 2008; Hattori et al., 2009; Gamuyao et al., 2013). Abundant quantitative variation for panicle traits is known to exist both within and between subpopulations (Ikeda et al., 2004; Zhao et al., 2011). The International Rice Research Institute (IRRI) Gene Bank reports panicles with numerous branching phenotypes and lengths ranging from 10 to 43 cm (Ruaraidh Sackville Hammilton, IRRI, Philippines, personal communication). Without a quick and accurate method of measuring panicle morphology, how can we hope to identify and leverage the genetic mechanisms underlying this spectrum of diversity? Is there an optimal panicle architecture to enhance yield in *O. sativa*, and is it specific to a particular subpopulation or environment? If so, can we maximize productivity and resource use efficiency by selectively tweaking reproductive development?

The Green Revolution gene *SEMIDWARF1* (*SD1*), which increased harvest index by reducing the ratio of vegetative to reproductive tissue, is an excellent example of how breeding for changes in rice plant architecture can significantly improve yield potential (Hedden, 2003; Asano et al., 2011). However, despite a sizable body of molecular and physiological research on rice panicle development (Wang and Li, 2011; Yoshida and Nagato, 2011), breeding for panicle architecture *per se* has had little success with regard to improving yield. This may be due in part

to phenotyping gaps within rice breeding pipelines, which often use composite traits like "panicle length" that are not always developmentally or botanically derived.

Numerous studies have used 2D image-based phenotyping to measure rice root architecture (Famoso et al., 2010; Iyer-Pascuzzi et al., 2010; Clark et al., 2011; Famoso et al., 2011) and grain traits (Tanabata et al., 2012). Two software packages have previously been developed to measure panicle-related traits, PASTAR/PASTA Viewer and P-TRAP (Ikeda et al., 2010; AL-Tam et al., 2013). PASTAR/PASTA Viewer extracts length and count traits from panicle images by identifying and connecting the top and bottom termini of branches (Ikeda et al., 2010). This platform was tested in a quantitative trait loci (QTL) mapping study (Ikeda et al., 2010), yet is under license and not freely available. The second platform P-TRAP is open-source, yet has not been validated in a genetic mapping study and was released after being tested on only 26 images (AL-Tam et al., 2013). P-TRAP converts a panicle image into a mathematical graph and treats branch junctions as vertices (AL-Tam et al., 2013). Although P-TRAP can also quantify several grain traits directly from a panicle image, its skeleton calculation is fundamentally similar to PASTAR/PASTA Viewer; vertices are connected using a series of straight line segments (AL-Tam et al., 2013). Straight-line skeletons can only approximate the true nature of curved axes like panicle branches, limiting the resolution of length phenotypes. Manually arranging a panicle before imaging can partially overcome this limitation, yet significantly increases time cost for trait evaluation. Additionally, neither PASTAR/PASTA Viewer nor P-TRAP can automatically eliminate awns, hair-like extensions on rice spikelets that can be misconstrued as branches.

The objectives of this study were: 1) to establish a flexible phenotyping methodology and platform capable of quantitatively measuring primary panicle architecture in rice; 2) to determine whether increased phenotyping resolution improved our ability to identify panicle-related QTLs in a population of recombinant inbred lines (RILs); and 3) to test our approach on a diverse set of rice varieties grown under different planting regimens in a controlled environment.

4.2 Results

Development and Implementation of the Phenotyping Platform PANorama

Within plants, meristem transition points are developmentally important and should be captured and separated functionally. In a rice panicle, meristem transitions along the main axis are relatively obvious because of the presence of trace bracts. The traditional breeding phenotype panicle length (PL) not only ignores meristem transitions, but also lumps in the length of the rachis axis with that of the terminal primary branch (Figure 4.1) (Ikeda et al., 2004; Ikeda et al., 2010). Following this logic, measuring rachis length using the IM abortion point is also not ideal, as the rachis can be considered a compound phenotype. Every point at which a primary branch forms represents a functional meristem transition during PI, from the IM to a PBM. Thus, count measurements are representative of the number of times a meristem transition occurs in development, and when multiple PBM transitions occur closely together they should be clustered into internodes along the rachis axis (Figure 4.1). Accordingly, length measurements may represent the amount of time between each transition and/or hormonal expansion after PI is completed, while width measurements may relate to physiological traits such as vascular architecture.

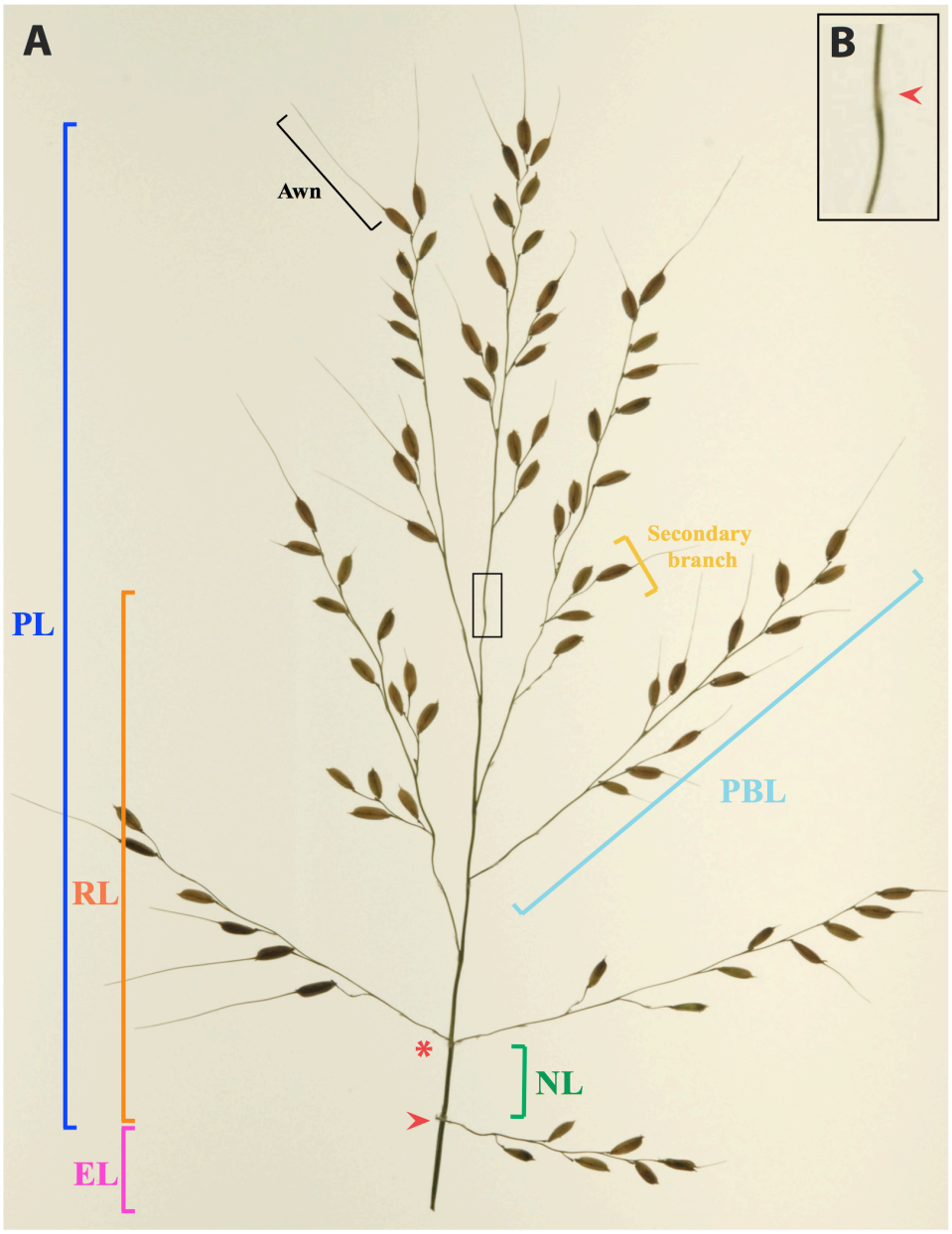


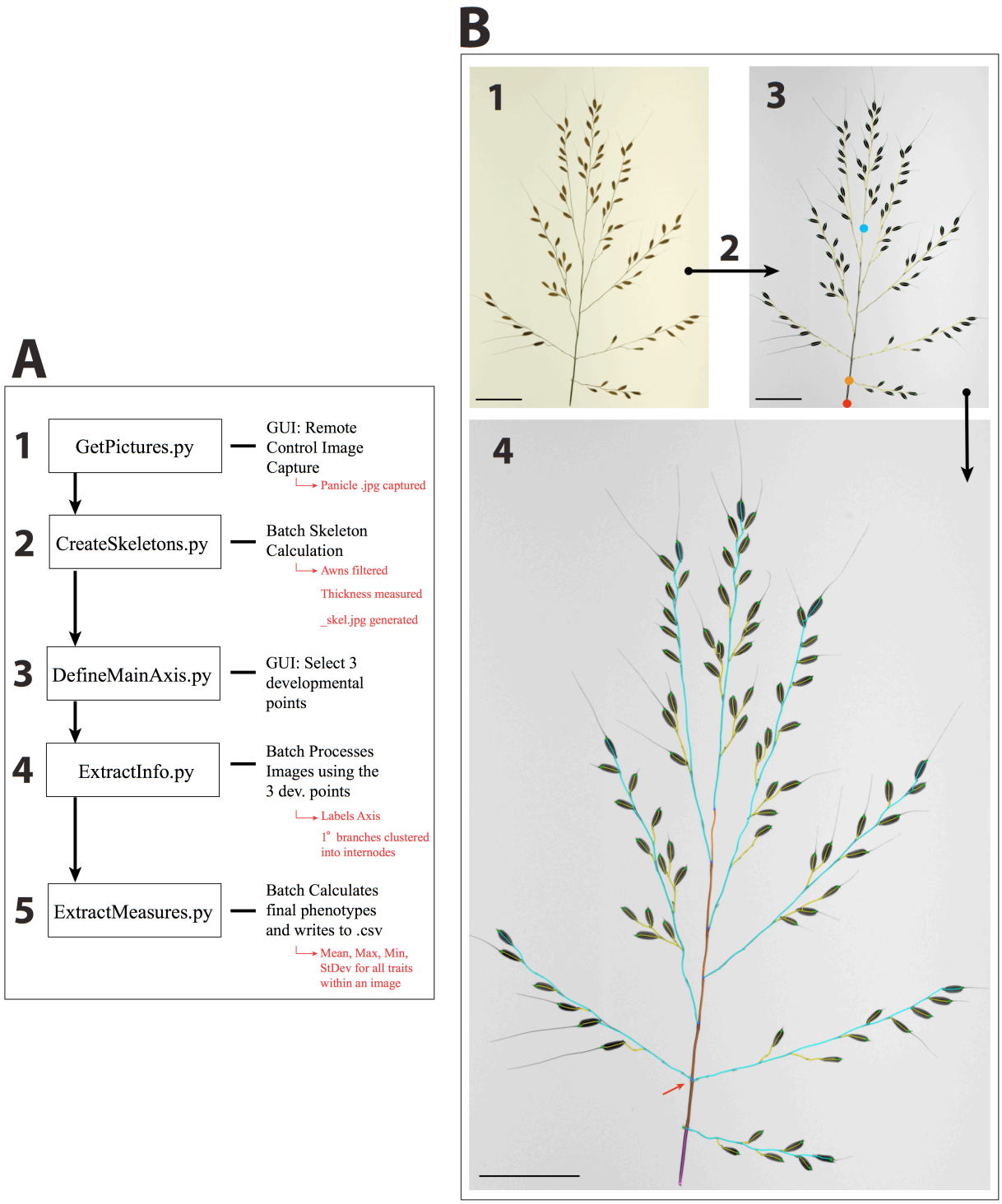
Figure 4.1. Inflorescence architecture in rice. A, Rice panicles are characterized by hierarchical transitions between meristems. Plant breeders often use the composite phenotype PL, even though it combines the length of the rachis axis (RL) with that of the terminal primary branch. Thus, several structural components should be accounted for when using image analysis to measure rice panicles. The panicle notch (red arrowhead) and IM abortion point (red arrowhead in B) denote the beginning and end of IM division, respectively. Multiple IM-to-branch meristem transitions can occur at a single internode point (red asterisk), and the distance between two such points can thus be defined as the NL. Measuring PBL is also important but can be complicated by the presence of awns and secondary branches. The vegetative phenotype EL is calculated as the distance from the flag leaf ligule to the panicle notch and is reflective of how far the panicle emerges from the leaf sheath during heading.

When using 2D image analysis and skeletonization techniques to accurately measure length, width, and count traits, an ideal skeleton should be a thinner representation of actual panicle axes. Previous studies have computed skeletons based on the exact morphological erosions of shapes using the image foresting transform (IFT) algorithm (Falcao et al., 2002; Falcão et al., 2004). Within this study, we incorporated this skeletonization technique into a new plant phenotyping pipeline, PANorama. PANorama is a collection of Python scripts that photograph, skeletonize, and extract multiple phenotypes from images. Its source code is written in C and the scripts essentially call the stand-alone C programs. There are five major steps in the PANorama pipeline, which are performed using a combination of automated batch processing and user-directed input via graphical user interfaces (GUIs) (Figure 4.2). Scripts are executed at different levels within a structured directory (Figure S1; Supplementary Videos), and multiple parameters that control the implementation of the pipeline are stored in one file, iftPANorama.h (Table S1). Thus, a user can edit iftPANorama.h based on the qualities of images in a directory and recompile the program before running the pipeline. The following sub-sections discuss how the parameters in iftPANorama.h have been calibrated across the pipeline for use on rice panicle images.

Image acquisition (GetPictures.py)

Image capture represents a large source of human error and directly limits the speed and accuracy of high-throughput phenotyping platforms. Regardless of the software platform used (Ikeda et al., 2010; AL-Tam et al., 2013), two-dimensional image analysis requires panicles to be positioned within a camera's field of view so that branches are non-overlapping. To reduce the need for post-image capture processing, the GetPictures.py GUI was incorporated into the

Figure 4.2. PANorama pipeline. PANorama generates a skeletonized panicle image based on the exact morphology of the panicle. A, Image capture, skeletonization, processing, and phenotype extraction are controlled by five scripts. B, Panicle images from different stages of the pipeline. Numbers correspond to the steps listed in A. 1, A panicle image captured and saved using GetPictures.py. 2, Execution of CreateSkeletons.py batch generates skeletons for every panicle image. 3, A _skel.jpg version of the original panicle image generated by CreateSkeletons.py. The skeleton is depicted in yellow, superimposed over the panicle axes. Using the DefineMainAxis.py GUI, three developmental checkpoints are selected using a mouse: EP (red circle), IMAP (orange circle), and FMAP (blue circle). 4, The _skel.jpg image after ExtractInfo.py has been executed. Using the three developmental checkpoints selected, ExtractInfo.py labels the skeleton and defines axes: exsertion axis (pink), RL (orange), primary branches (blue), and secondary and tertiary axes (yellow). Branches are collapsed into internodes during this step (red arrow). Bars = 500 pixels.


A
1
GetPictures.py
GUI: Remote Control Image Capture
Panicle .jpg captured
2
CreateSkeletons.py
Batch Skeleton Calculation
Awns filtered
Thickness measured
_skel.jpg generated
3
DefineMainAxis.py
GUI: Select 3 developmental points
4
ExtractInfo.py
Batch Processes Images using the 3 dev. points
Labels Axis
1° branches clustered into internodes
5
ExtractMeasures.py
Batch Calculates final phenotypes and writes to .csv
Mean, Max, Min, StDev for all traits within an image
B
1
2
3
4

PANorama pipeline. GetPictures.py interacts with the open access software gPhoto2, allowing a user to remotely control image capture via his/her computer. All imaging in this study was conducted using a fixed camera mount and a light box, so that images are of the same exposure and scale (see Materials and Methods). After image capture, GetPictures.py generates a preview of a panicle's skeleton so that the user can correct misarranged branches and retake the photo. Images are saved as a .jpg file within a directory titled "originals," and image names are stored in a .txt file *imagefiles.txt* (Figure S1). Every other script in the pipeline uses *imagefiles.txt* to access images stored within "originals" (Video S1).

Skeletonization (CreateSkeletons.py)

The second step of the pipeline, CreateSkeletons.py, uses a combination of image processing operators to segment the panicle from the image background, remove awns, and calculate a skeleton for each image within the "originals" directory. For every original image, two new files are generated with the same root name. The first is a _skel.jpg file, containing a gray scale version of the original image with the final skeleton superimposed on the panicle (Figure 4.2B); the second is a .inf extension file containing information associated with skeleton calculation. Both of these files are stored in a parallel directory named "skeletons" (Figure S1). The entire process is fast and takes ~10 seconds for a 2592 x 3872 pixel image (Video S2).

Given that some panicles have awns, we incorporated a pre-processing step into CreateSkeletons.py that is automatically performed before final skeleton calculation (Figure 4.3A-4.3E). First, the Euclidean distance transform (Falcao et al., 2002) is used to compute an exact erosion of 0.032cm. The erosion eliminates awns, but also breaks the panicle mask into

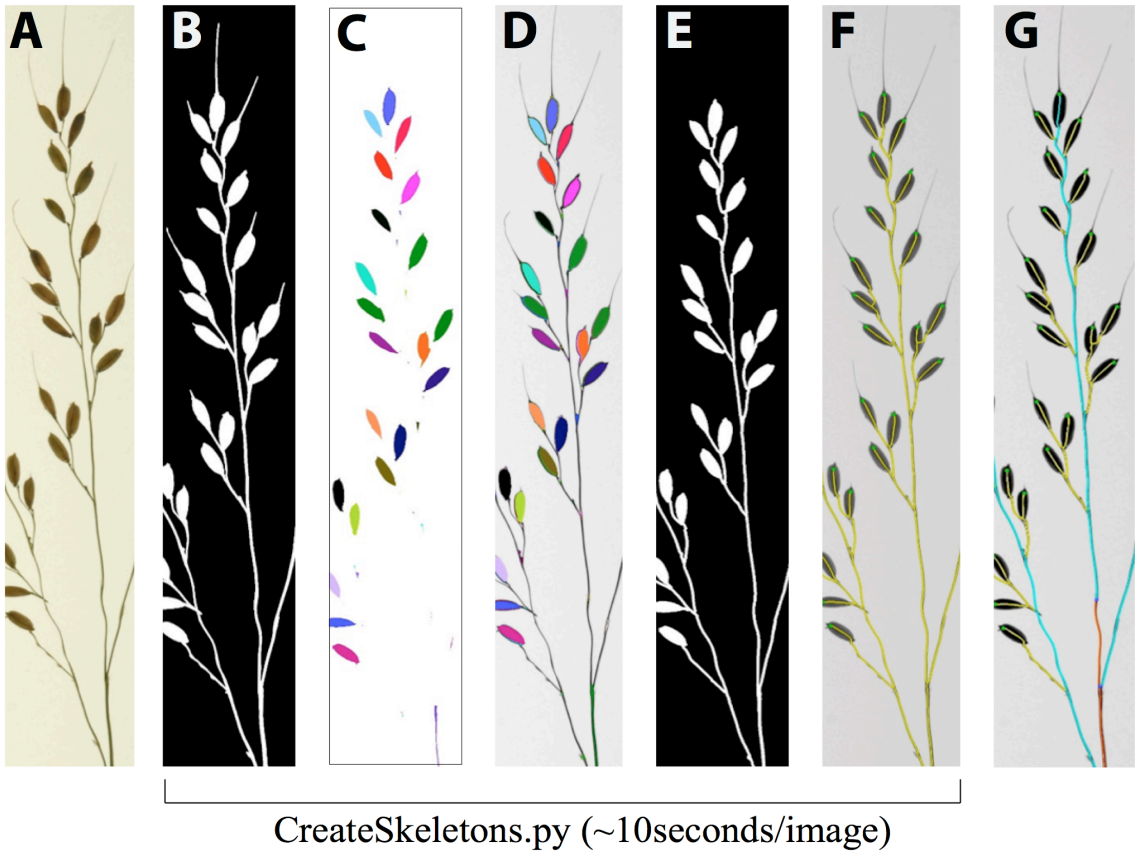
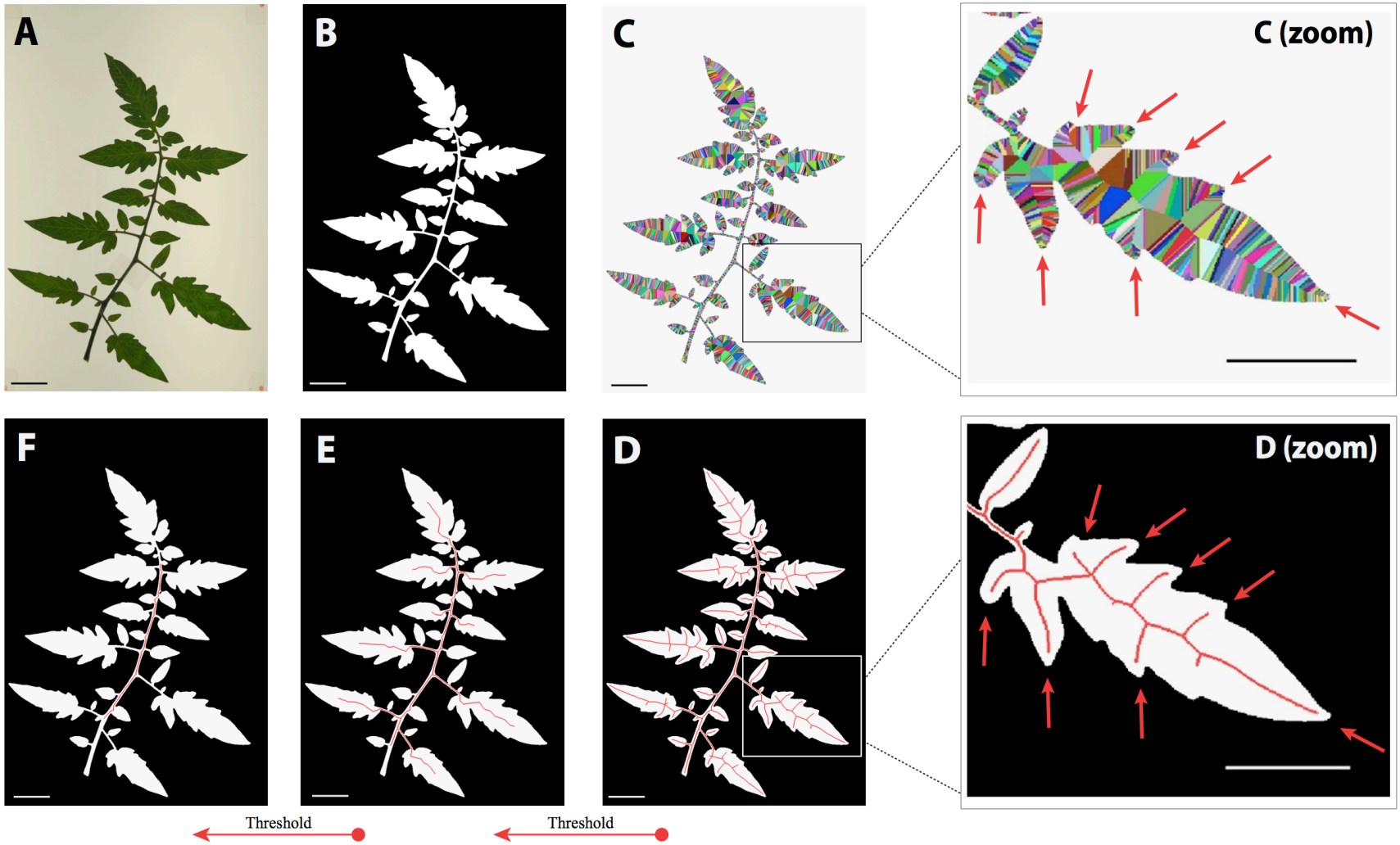


Figure 4.3. Schematic depicting awn removal using exact morphological erosions. A, Raw image of a panicle branch. B, Initial binary image after segmentation. C, Erosion of the image at 0.032 cm using the Euclidean distance transform. The binary components of the image object that remain after erosion are labeled with distinct colors. D, Binary components are superimposed over the original panicle image and reconnected by the darkest path (i.e. branches and rachis axis). E, After path reconnection, the awnless binary image is used to calculate the final skeleton. F, Panicle skeleton (yellow) superimposed over axes. All segmentation and skeletonization steps (B–F) are controlled by the CreateSkeletons.py script. G, Final panicle skeleton after the use of DefineMainAxis.py and ExtractInfo.py.

several connected components. Because the pixels from panicle axes and adjacent connected components (i.e. seeds) appear dark in the brightness image, these components are labeled (Figure 4.3C) and reconnected by paths that contain a minimum sum of pixel values between each pair of distinctly labeled components (Figure 4.3D). A final binary mask for skeletonization is computed by exact dilation at 0.032 cm, followed by a morphological closing of holes. Although this generates a mask with a single contour (Figure 4.3E), the holes within panicle branches are still used to define the final skeleton (Figure S2). This is important when calculating skeletons for panicles in which primary branches have not been arranged to have open secondary or tertiary branches (next section). Euclidean distance transform (exact erosions), path reconnection, and closing of holes are all computed by the Image Foresting Transform algorithm (Falcão et al., 2004) during execution of CreateSkeletons.py.

Figure 4.4 depicts skeleton calculation using a tomato leaf, which is wider and easier for visualization than a rice panicle. First, a RGB color image (Figure 4.4A) is segmented and thresholded to generate a binary image (Figure 4.4B). This is done using morphological erosion with a circular structuring element of radius 0.083mm on its brightness values to enhance the dark pixels that represent a panicle, and a fixed factor of 1.2 on the Otsu's optimum threshold. To eliminate spurious structures that are not physically connected to panicles (dust particles or detached seeds), the resulting binary mask contains only the largest connected component (Figure 4.4B). The skeletonization process proposed in (Falcao et al., 2002) uses the Euclidean distance transform to propagate the closest contour pixel (called root) to each internal pixel of the panicle shape. This process can be visualized as wavefronts of distinct colors, which propagate from each contour pixel and meet one another in the center of the image at pixels

Figure 4.4. Skeleton calculation using the image foresting transform algorithm. Using an RGB color image of a tomato leaf (A), a binary mask is generated from the largest eight-connected component (B). A Voronoi discrete map (C) is calculated using the Euclidean distance transform. This process is visualized as wave fronts of distinct colors, which propagate inward from each contour pixel (C). When pixels meet in the center of the shape, they generate central contours (C zoom, arrows). These contours are used to define the skeleton of an image object (D zoom, arrows). By weighting the importance of pixels with respect to their positions along the length of the skeleton (i.e. along the scales), a skeleton can be thresholded using the exact morphology of the shape (E and F). For rice images, this scale parameter is set at 0.1%, as observed in D.


A
B
C
C (zoom)
F
E
D
D (zoom)
Threshold
Threshold

equidistant to the contour (Figure 4.4C, arrows). For each internal pixel, the method computes the maximum among the geodesic distances along the contour between the root of the pixel and the root of its 4 neighbors. These measurements are assigned to each internal pixel and represent multi-scale skeletons, which can be simplified by increasing thresholds on the scale (Supplementary Methods). The generated skeletons are all one-pixel-wide and fully connected (Figures 4.4D-4.4F), which makes it very simple to identify nodes and extract accurate length measurements from a panicle. Additionally, skeletons are in the center of the shape at meeting points between distinct wavefronts (Figure 4.4D, arrows). Because of this, distances to the contour can be doubled to accurately estimate thickness measurements along panicle segments (see Supplementary Methods). For rice images, we set the scale threshold at 0.1% of the contour length using the SKELTHRES parameter in iftPANorama.h.

Image Orientation (DefineMainAxis.py)

After obtaining accurate skeletons from panicle images, it is necessary to define the orientation of an image prior to phenotype extraction. Executing the DefineMainAxis.py script opens a GUI that allows a user to load individual _skel.jpg files. The user can select, move, and save three points on top of a panicle skeleton: Extrusion Point (EP), Initial Main Axis Point (IMAP), and Final Main Axis Point (FMAP) (Figure 4.2). For rice panicles, we identify EP as the point at which the flag leaf ligule attaches, while IMAP and FMAP are selected based on the presence of the panicle notch and the IM abortion point, respectively (Figure 1). Manual selection of these points is essential, as their position is variable for every image. For each _skel.jpg analyzed, the selection points are stored within a .txt file of the same root name within the “skeletons”

directory (Figure S1; Video S2). This file is not edited if a user needs to recalculate a skeleton using CreateSkeletons.py, so that the three points only have to be selected once for each image.

Phenotype Extraction (ExtractInfo.py and ExtractMeasures.py)

Phenotypes are automatically calculated using two scripts: ExtractInfo.py and ExtractMeasures.py. ExtractInfo.py assigns and labels the orientation of panicle axes for each _skel.jpg image, using the three points identified with DefineMainAxis.py. After running ExtractInfo.py, the _skel.jpg image depicts the final, labeled skeleton (Figure 4.2B). Several preset thresholds stored in the iftPANorama.h file are important for this step, as ExtractInfo.py filters some phenotypic traits before final measurements are calculated (Figure 4.2). For example, branches are excluded from final measurement calculation if they are below a certain length, and multiple primary branches are collapsed into single internodes if they attach to the rachis axis within a certain distance of one another (see Supplementary Methods; Table S1; Video S3).

The final thresholded and labeled _skel.jpg files are used by ExtractMeasures.py to calculate phenotypes for primary panicle architecture. ExtractMeasures.py operates on every image across multiple directories and takes ~5-7 seconds per image (Figure S1; Video S3). Table 4.1 lists phenotypes and how they are generally defined within the pipeline. Details regarding the algorithms used are stored within Supplementary Methods. ExtractMeasures.py outputs two .csv files; the first contains mean, maximum, minimum, and standard deviations for phenotypes within each image (*PanicleMeasures.csv*), while the second contains the raw phenotypic

Table 4.1. Phenotype calculations in PANorama. The following phenotypes are written to a .csv file after running ExtractMeasures.py, using the PANorama phenotype designation. For simplicity, we refer to rice phenotypes within the rest of the chapter using QTL abbreviations. EP = Exsertion Point, IMAP = Initial Main Axis Point, FMAP = Final Main Axis Point (Figure 4.1)

PANorama phenotype designation	Calculation	QTL abbr.
LengthOfExtrusion(cm)	The distance between EP and IMAP. In rice, these points are the flag leaf ligule attachment point and the panicle notch, respectively.	EL
LengthOfMainAxis(cm)	The distance between IMAP and FMAP. In rice, these points are the panicle notch and inflorescence abortion point, respectively.	RL
InflorescenceLength(cm)	The distance between the IMAP and the tip of the longest terminal branch. In rice, IMAP is selected at the panicle notch.	PL
NumberOfPrimaryBranches	A primary branch is any distinct skeleton segment that attaches (i.e. shares a neighboring pixel with the main axis) between IMAP and FMAP. The MINBRANCSZ parameter defines the minimum length requirement of a primary branch.	PBN
MeanLengthOfPrimaryBranches(cm)	The distance from the first pixel in a primary branch that borders the rachis skeleton, to the eight-neighbor pixel at the tip of a terminal seed. PBL is the average of all primary branches within a panicle.	PBL
MaxLengthOfPrimaryBranches(cm)	The maximum length of a primary branch within a panicle.	xPBL
MinLengthOfPrimaryBranches(cm)	The minimum length of a primary branch within a panicle.	nPBL
StdevLengthOfPrimaryBranches(cm)	The standard deviation of PBL.	
NumberOfNodesOnMainAxis	Nodes are defined as anywhere a meristem transition has occurred. The pixels at which primary branches attach are considered nodes. If a branch forms at the IMAP and FMAP, they are also treated as nodes. Branches that attach within a specified distance of one another are automatically collapsed into a single internode point using the MINJCTDIST parameter, and the pixel closest to the geometric center of the cluster is considered the true internode point. NN is the total number of nodes in an image.	NN
MeanNumberOfBranchesPerNode	The average number of branches per node for all nodes within a panicle. Branches are collapsed into single nodes using the MINJCTDIST parameter, and the pixel	BpN

	closest to the geometric center of the cluster is considered the true node.	
MaxNumberOfBranchesPerNode	The maximum number of branches to occur at a single node within a panicle.	xBpN
StdevNumberOfBranchesPerNode	The standard deviation of BpN.	
MeanLengthBetweenNodes(cm)	The distance between two nodes, after collapsing of primary branches. NL is the average length of all internodes in an image.	NL
MaxLengthBetweenNodes(cm)	The maximum distance between two nodes.	xNL
MinLengthBetweenNodes(cm)	The minimum distance between two nodes.	nNL
StdevLengthBetweenNodes(cm)	The standard deviation NL.	
MeanThicknessOfMainAxis(cm)	Thickness is calculated as 2x the distance between a skeleton point and its closest pixel on the image contour. Distances are obtained directly from the Euclidean distance map of the panicle mask during the skeletonization step. TR is the average of all thickness measures between IMAP and FMAP.	TR
MaxThicknessOfMainAxis(cm)	The maximum thickness between the IMAP and the FMAP.	xTR
StdevThicknessOfMainAxis(cm)	The standard deviation of TR.	
MeanThicknessOfExtrusion(cm)	Thickness is calculated as 2x the distance between a skeleton point and its closest pixel on the image contour. Distances are obtained directly from the Euclidean distance map of the panicle mask during the skeletonization step. TE is the average of all thickness measures between EP and IMAP.	TE
MaxThicknessOfExtrusion(cm)	The maximum of all thickness measures between EP and IMAP.	xTE
MinThicknessOfExtrusion(cm)	The minimum thickness measurement between EP and IMAP.	nTE
StdevThicknessOfExtrusion(cm)	The standard deviation of TE.	

measures (*AllPanicleMeasures.csv*). Length and width measurements are derived using a pixel to length conversion, which is dependent on the scale of the image (see Materials and Methods).

Error Correction (CorrectBranches.py and RebuildImageFiles.py)

Two supplementary scripts are included in PANorama for error correction. The RebuildImageFiles.py script regenerates the *imagefiles.txt* file that lists the name of every image stored within "originals." Rebuilding *imagefiles.txt* is necessary if a user edits the "originals" directory by adding or removing images, and wishes to re-analyze the edited directory using the other scripts in the pipeline (Video S1). Executing CorrectBranches.py opens a GUI that allows users to add, remove, or shorten primary branches that were incorrectly arranged during image capture. The edits are saved directly within the .inf file of the image being edited (Video S3).

Quality testing of the PANorama pipeline

Comparison of Manual Measurements, PANorama, and P-TRAP

To compare manual measures to PANorama and P-TRAP measures, we used GetPictures.py to photograph 30 panicles from the parents of the RIL mapping population in this study (IR64 and Azucena). Each panicle was photographed two times using different arrangements within an image (N=60 images); once with secondary and tertiary branches closed, and once with all branching arranged openly (Figure S2). Manual measurements were taken by hand (see Materials and Methods). To test awn filtering on primary branches, PANorama skeletons were calculated with the AWNSTHICK parameter in fitPANorama.h turned on and off (on=0.032cm, off=0.0cm). Panicle images were also processed using the open-source software P-TRAP (AL-

Tam et al., 2013). For PANorama and P-TRAP, the panicle notch and IM abortion points were selected in order to extract phenotypes. However, skeletons were not corrected in any other way.

Table 4.2 reports squares of Pearson's Correlation Coefficients (r^2 values), mean absolute percentage deviations (MAPDs), and mean percentage errors (MPEs), between manually measured phenotypes and the phenotypes calculated by each of the software packages. Measuring opened panicle branches in PANorama decreased MAPD from manual measurements for PBL by ~1%, when compared to measuring closed panicle branches. Additionally, MAPDs were always lower for PBL when the awn filter was turned on (AWNSTHICK=0.032cm), regardless of whether branches were open or closed (Table 4.2). This suggests that awn filtering always increases accuracy of primary branch measures, even in varieties like IR64 and Azucena that have very small awns (Figure S2). The squares of correlation coefficients between PANorama phenotypes were high for every reported phenotype (Table S2), confirming that the AWNSTHICK parameter accurately filters awns without negatively impacting other phenotypes.

When compared to manual measurements, PANorama phenotypes always showed higher r^2 values, lower MAPDs, and MPEs closer to zero than P-TRAP. Additionally, PANorama required much less computational time to generate skeletons and calculate phenotypes (Table 4.2). When comparing PANorama and P-TRAP skeletons visually, we found that P-TRAP was less accurate at identifying curved axes (Figure S3A). This could partially explain why P-TRAP misrepresented RL and NN, resulting in high MAPDs and strongly negative MPEs for both traits (Table 4.2). P-TRAP also could not distinguish awns from branches (Figure S3B).

Table 4.2. Comparison of manual, PANorama, and P-TRAP measurements. Squares of Pearson's Correlation Coefficient (r^2 values), mean absolute percentage deviations (MAPDs), and mean percentage errors (MPEs) were calculated using 30 images, between manual measurements and phenotypes calculated by each software platform respectively. Total processing time for 60 images includes skeletonization and measurement calculation, but excludes selection of meristem transition points.

	PANorama				P-TRAP		
Trait	Parameter	r^2	MAPD	MPE	r^2	MAPD	MPE
RL	NA	0.998	0.78	-0.004	0.06	59.51	-26.20
PBN	NA	0.99	0.42	0.004	0.77	3.16	-2.25
NN	NA	0.86	5.76	-0.050	0.42	28.26	-20.10
BpN	NA	0.83	6.35	0.065	NA	NA	NA
PBL (open)	AWNSTHICK (0.032cm)	0.996	1.42	0.002	NA	NA	NA
PBL (closed)	AWNSTHICK (0.032cm)	0.998	2.54	-0.019	NA	NA	NA
PBL (open)	AWNSTHICK (0.0cm)	0.996	2.59	-0.021	0.93	9.55	-3.26
PBL (closed)	AWNSTHICK (0.0cm)	0.996	3.69	-0.032	0.90	6.48	2.85
Processing time	10 minutes, 30 seconds				1 hour, 10 minutes		

PANorama in Other Species

In order to test the applicability of PANorama in other organisms, we photographed and analyzed the following plant structures: a male inflorescence from corn (*Zea mays*); a compound leaf and an inflorescence from tomato (*Solanum lycopersicum*); and an entire Arabidopsis plant (*Arabidopsis thaliana*) (Figure 4.5). With minimal tweaking to the parameters within iftPANorama.h, we were able to calculate phenotypes across these differently shaped image objects. The three points selected in DefineMainAxis.py were chosen in order to measure different biological structures within each organism (see Materials and Methods; Table S3).

Validation of PANorama Phenotypes in *Oryza sativa*

QTL Analysis

To test the biological validity of PANorama phenotypes, we performed a QTL analysis using a population of RILs derived from a wide cross between IR64 (*Indica*) and Azucena (*Japonica*). In any phenotype mapping study, there are tradeoffs between the number of replicates, number of traits, and the resolution of measurements that are feasible to obtain. Because the MAPDs of open versus closed primary branches differed by ~1% (Table 4.2), we phenotyped primary architecture in high replication (N=1537 images) and forewent manual arrangement of higher branching (Figure S2A). QTL mapping was performed following the methods described in (Spindel et al., 2013), using 30,984 single nucleotide polymorphisms (SNPs) generated via genotyping by sequencing (GBS). Numerous QTL were detected throughout the genome for length, width, and count measurements (Table 4.3; Table S4). We also detected several QTL for two reproductive traits that were measured without using PANorama; days to heading (DTH) and panicle number (PN) (Table S4; see Materials and Methods). Several regions of the genome

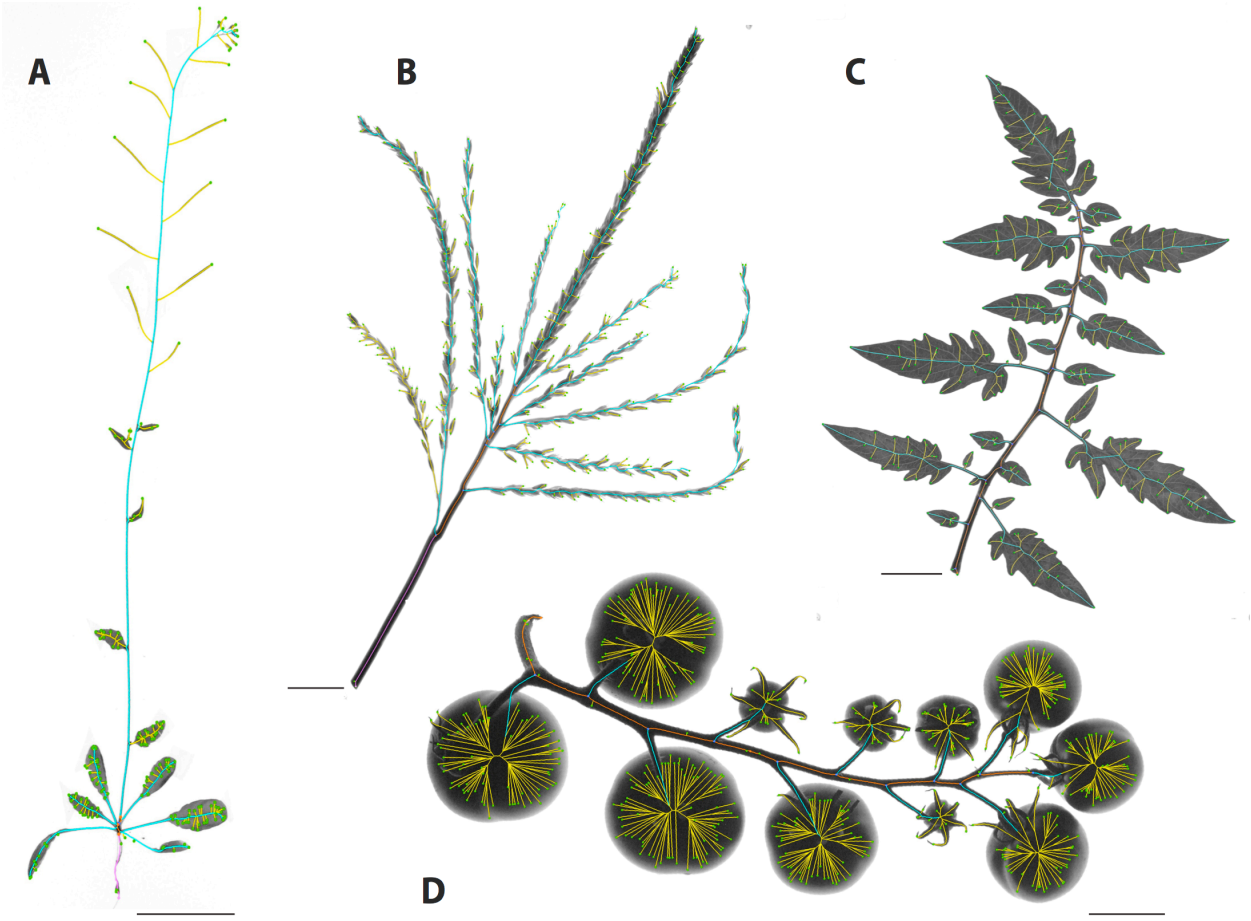


Figure 4.5. PANorama in other species. With minimal adjustments to parameters stored in the iftPANorama.h file, PANorama can be easily adapted to measure structures from other organisms, including total plant architecture in Arabidopsis (A), the male inflorescence of maize (B), and the compound leaf (C) and inflorescence (D) of tomato. Phenotypic values are listed in Supplemental Table S3. Bars = 500 pixels.

Table 4.3. Panicle trait QTL identified using PANorama phenotypes. Total PVE represents the percent variance explained by a multi-QTL model.

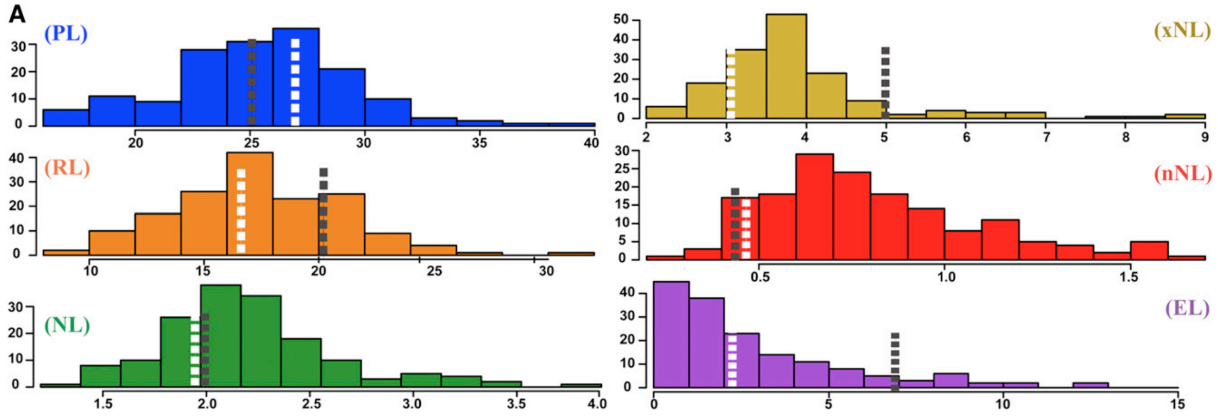
Trait	Trait Type	Chromosome	Total PVE	# of QTL
EL	Length	1, 4, 6, 8	49.77	5
RL	Length	1, 8	32.50	2
PL	Length	1	17.26	1
NL	Length	1, 5, 6, 9, 10	56.15	6
xNL	Length	1, 5	23.25	2
nNL	Length	1, 3, 9, 11	42.39	4
PBL	Length	1, 3, 4, 8, 11	40.26	5
xPBL	Length	4	10.80	1
nPBL	Length	6	9.99	1
PBN	Count	3, 4	35.86	3
NN	Count	3, 8	25.19	2
BpN	Count	1, 4	28.76	2
xBpN	Count	1, 3, 4, 7	44.30	4
TR	Width	1, 4, 7	30.86	3
xTR	Width	4	17.02	1
TE	Width	4	10.86	1
xTE	Width	1, 4	20.01	2
nTE	Width	2, 8	22.65	2
Total number of PANorama QTL identified				47

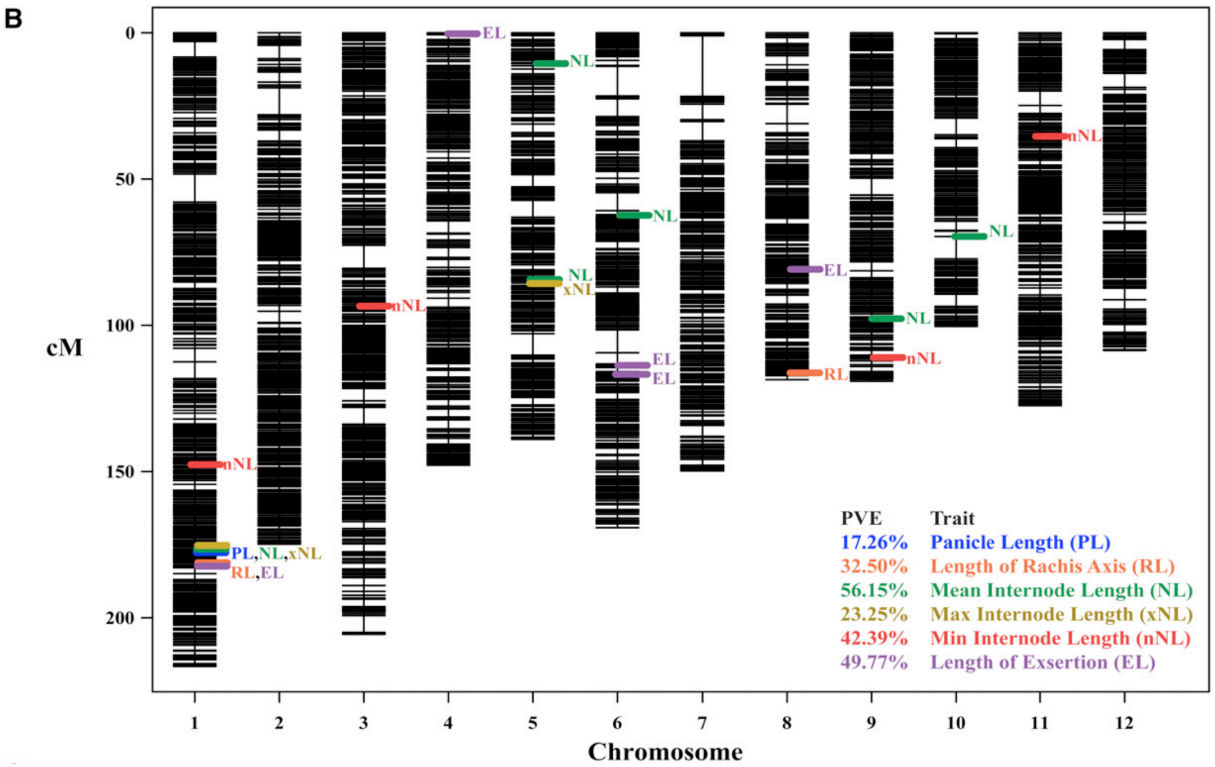
showed clusters of QTL for PANorama panicle traits and/or these independently measured reproductive phenotypes (Table S4). Figure 4.6 and Table 4.4 summarize results for six traits, including one vegetative phenotype exsertion length (EL) and multiple phenotypes that represent components of the composite trait panicle length (PL). All of these phenotypes display transgressive variation in the RIL population compared to the parents (Figure 4.6A). Notably, despite using PANorama, panicle length had a single QTL on chromosome 1 explaining 17.26% of the phenotypic variation (Table 4.4). By dividing panicle length into the rachis length, internode length, maximum internode length, and minimum internode length traits, we identified 16 additional QTL at 11 distinct locations throughout the genome and increased the total percent variance explained (Figure 4.6B; Table 4.3). Of particular interest, nearly every length phenotype within this study, vegetative and reproductive, mapped to a large region on the distal end of chromosome 1 (Table S4). Figure 4.6C depicts the LOD profiles for a few of these traits from ~140 cM to the end of chromosome. Notably, peak markers for panicle length, maximum internode length, and average internode length show increasing LOD scores of 6.23, 9.67, and 12.17, respectively. The largest peak markers for rachis length and exsertion length are shifted roughly 3 cM downstream, with LOD scores of 7.82 and 14.45 respectively (Figure 4.6C; Table 4.4). These results demonstrate that biologically precise phenotyping using PANorama complements dense genetic marker data, improving the resolution of QTL peaks at a single genomic locus.

Comparison of Panicle Traits Across Genetic Lines

To establish the extent of morphological variation in rice panicle architecture, both within a variety and across different genetic backgrounds, we grew a small panel of eight lines in varying

Figure 4.6. PANorama improved our ability to measure and map for quantitative panicle traits. A, Frequency distributions of phenotypic line means for six traits, calculated using 158 RILs and the two parents of the mapping population. The total number of panicle images was 1,522. The means for IR64 and Azucena are labeled using black or white dashed lines, respectively. The x axes depict length measurements (cm) for each phenotype. B, Genetic map of the 12 chromosomes in rice, constructed with 30,984 single-nucleotide polymorphisms. PVE indicates total PVE for all QTLs within a trait. C, QTL peaks at the distal end of chromosome 1, shown from approximately 140 cM to the end of the chromosome. The y-axis depicts the LOD score calculated for each marker across the region. QTL mapping was performed using Haley-Knott regression and 1,000 permutations to determine a significance threshold for each phenotype; the average threshold for all 6 phenotypes is 3.42 (not shown). LOD profiles show the refined QTL locations after forward selection and backward elimination were used to probe the model space. The peak marker for each trait is depicted using a vertical dashed line.




B
cM
0
50
100
150
200
EL
NL
nNL
NL
xNL
nNL
EL
EL
nNL
PL,NL,xNL
RL,EL
EL
RL
NL
nNL
NL
nNL
PVE
Trait
17.26%
Panicle Length (PL)
32.50%
Length of Rachis Axis (RL)
56.15%
Mean Internode Length (NL)
23.25%
Max Internode Length (xNL)
42.39%
Min Internode Length (nNL)
49.77%
Length of Exsertion (EL)
1
2
3
4
5
6
7
8
9
10
11
12
Chromosome

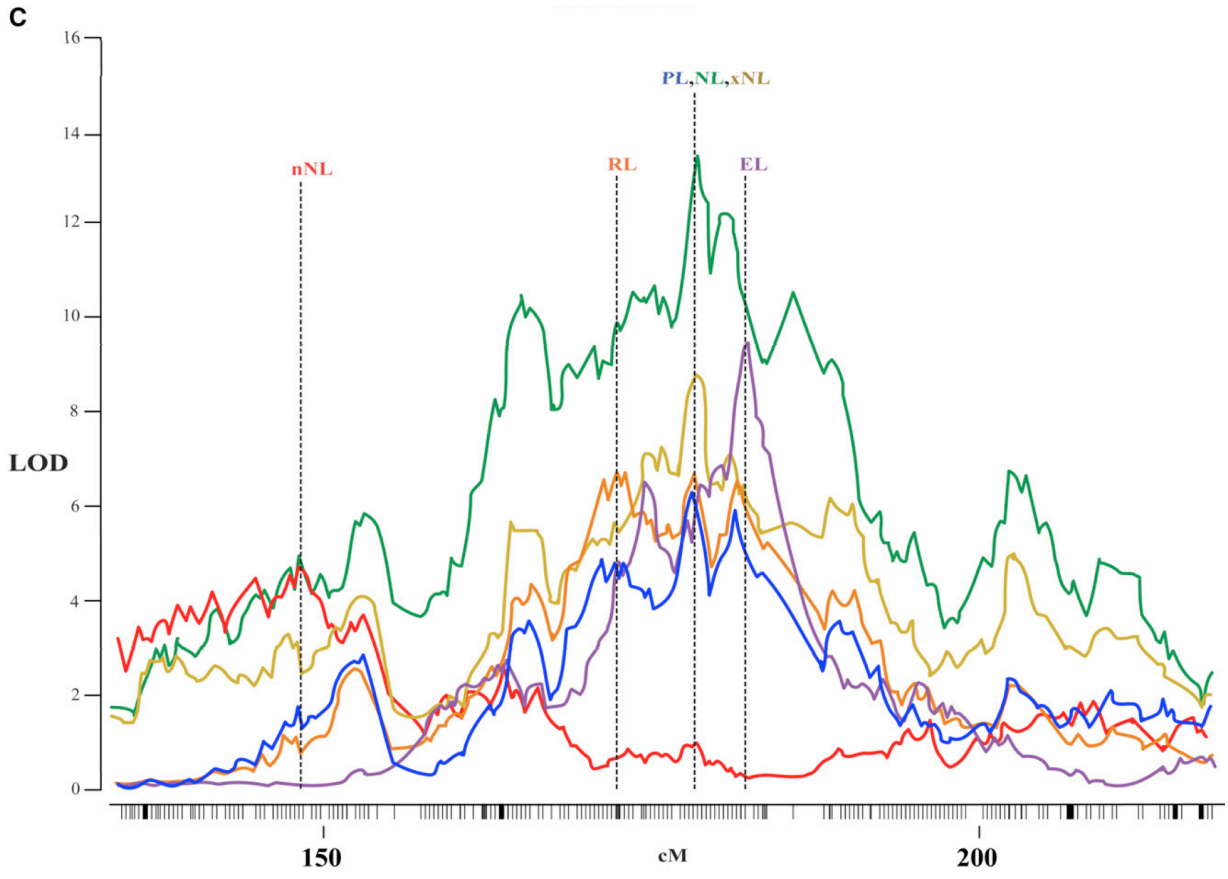

C
LOD
16
14
12
10
8
6
4
2
0
PL,NL,xNL
nNL
RL
EL
150
cM
200

Table 4.4. Dissection of panicle length into internode measurements using PANorama increases the number of QTL identified in an IR64 x Azucena RIL Population. Several peaks have been identified in previous studies (Pr. id.), but many are novel. Percent variance explained (PVE) was calculated for each peak. IR64 additive effects (Add. Eff.) are reported in units for the trait. Non-normal phenotypes were either transformed using the natural logarithm (*), or run with a nonparametric model ([np]).

Pheno	Chr	Peak(cM)	Peak Marker	L, R Marker	LOD	PVE	Add. Eff.	Pr. id.
nNL	1	148.44	S1_30057829	S1_26453459 S1_31817994	4.53	8.31	-0.05*	f
PL	1	178.17	S1_37484178	S1_35922333 S1_38642939	6.38	17.26	-1.66cm	a, b
NL	1	178.5	S1_37521762	S1_37370036 S1_38150186	13.53	21.7	-0.06*	a, b
xNL	1	178.5	S1_37521762	S1_37314452 S1_37838629	8.56	20.48	-0.55cm[np]	a, b
RL	1	172.11	S1_36198766	S1_35691398 S1_38679377	6.65	14.75	-1.58cm	a, b
EL	1	182.15	S1_38527519	S1_38116619 S1_38642939	9.38	16.15	-1.2cm[np]	a, c, e
nNL	3	94.16	S3_16076715	S3_15765689 S3_16438463	6.62	12.51	0.06*	
EL	4	2.32	S4_405005	S4_309716 S4_13379128	4.09	6.49	0.89cm	
NL	5	11.35	S5_1442539	S5_1247072 S5_1568034	5.93	8.45	-0.04*	
NL	5	85.13	S5_21207727	S5_20587858 S5_21705936	3.79	5.23	-0.05*	
xNL	5	85.13	S5_21207727	S5_1279922 S5_22515124	3.39	7.48	-0.34cm	
NL	6	62.63	S6_10319721	S6_8821624 S6_12046394	4.31	5.99	-0.03*	
EL	6	114.93	S6_22200268	S6_21980490 S6_22254152	3.97	6.29	-0.29cm[np]	
EL	6	116.58	S6_22260180	S6_22200268 S6_22410496	6.46	10.64	-0.57cm[np]	
EL	8	79.91	S8_22603837	S8_22484223 S8_23108392	6.5	10.64	-0.89cm[np]	a, c, e
RL	8	112.69	S8_27305892	S8_27213374 S8_28389376	7.29	16.32	-1.67cm	d
NL	9	96.74	S9_18987864	S9_16451344 S9_20174149	3.94	5.44	-0.03*	
nNL	9	110.9	S9_21397331	S9_20403572 S9_22503375	4.34	7.94	-0.04*	
NL	10	67.94	S10_19095358	S10_18448365 S10_19941941	5.47	7.74	0.03*	
RL	10	67.94	S10_19095358	S10_18437567 S10_21920802	3.69	5.06	1.08cm	
nNL	11	34.31	S11_5453415	S11_4362683 S11_5644994	7.33	14.02	0.04*	

(Hittalmani et al., 2003; Kobayashi et al., 2003)[b]; Li et al., 2003[c]; (Ikeda et al., 2010)[d]; Asano et al., 2011[e]; (Zhao et al., 2011)[f]

pot sizes under controlled environmental conditions (N=1,139 images). The eight lines represent the parents of four mapping populations currently available in the international rice community; they are representative four of the *O. sativa* subpopulations, and one wild relative: Azucena, Jefferson, and Moroberekan (*tropical japonica*); Nipponbare (*temperate japonica*); IR64 and CO39 (*indica*); Kasalath (*aus*); and *Oryza rufipogon* IRGC 105491 (wild). For many of the measurements, each line shows different means and distributions. The variation observed within varieties was much smaller than that observed between varieties, even when plants were grown under different pot sizes (Figure 4.7A-4.7D; Table S5). Although we detected subtle variation within a genetic line across pot sizes, differences were not statistically significant (Table S6).

4.3 Discussion

The rice inflorescence is a complex organ, developmentally and physiologically, and yet plant breeders have continued to measure panicles using very blunt tools. This is out of step with the major advances in sequencing and genotyping technology that have greatly increased dissection of complex traits by QTL mapping and/or genome-wide association studies (GWAS). The major goal of this study was to develop a platform and methodology that facilitates high-resolution phenotyping of rice panicles.

PANorama is an Accurate and Flexible Phenotyping Platform

For branched structures like the rice panicle, image analysis represents an accurate and high-throughput method to obtain multiple phenotypes simultaneously. However, phenotyping via image analysis has its own limitations, often requiring special preparation of plant tissue or time intensive, manual arrangement of samples before a photograph can be taken. Within this study,

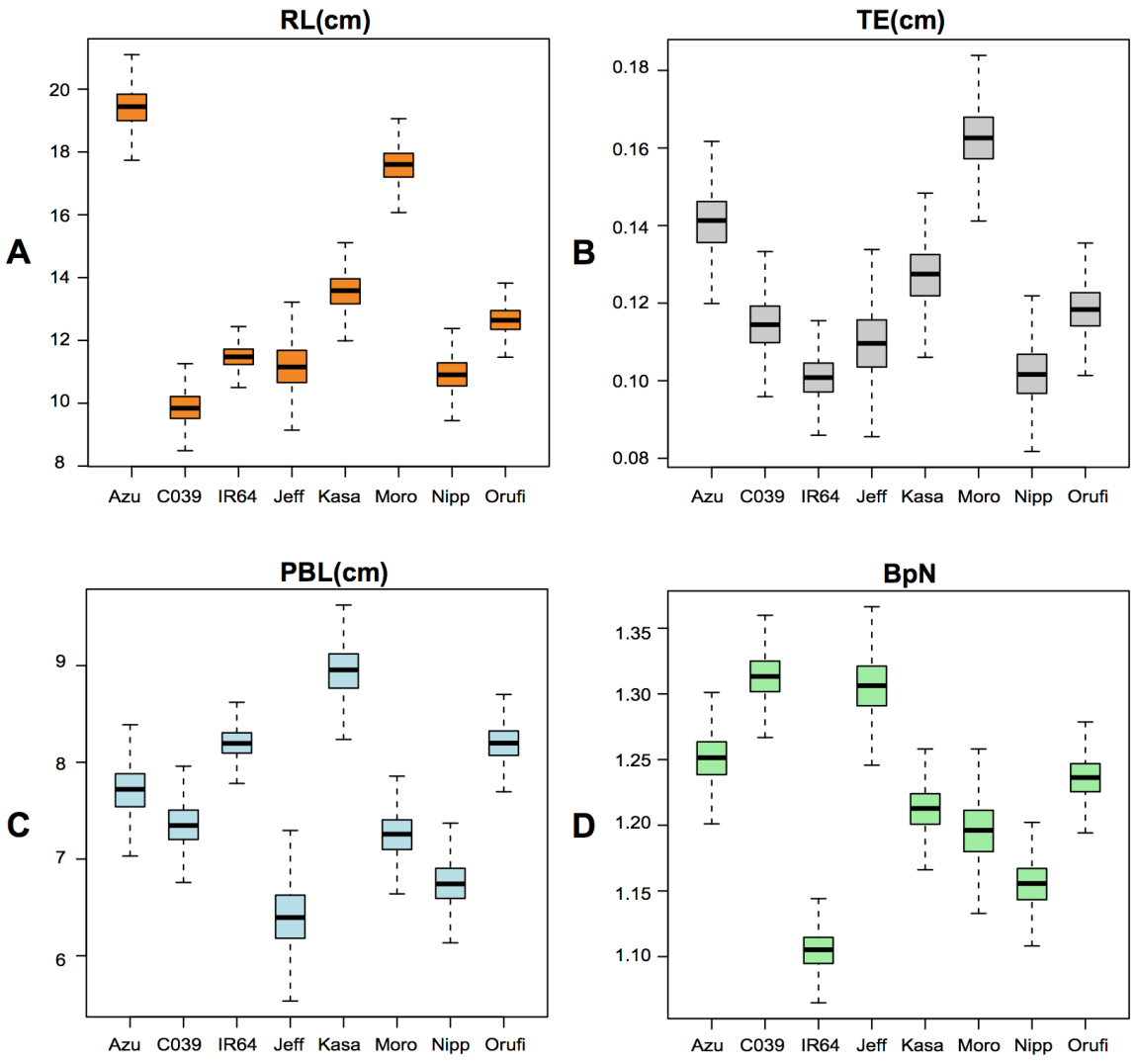


Figure 4.7. Quantitative variation in rice panicle traits. Phenotypic distributions are shown for seven distinct genetic lines of rice and one wild relative: Azucena (Azu), CO39, IR64, Jefferson (Jeff), Kasalath (Kasa), Moroberekan (Moro), Nipponbare (Nipp), and *O. rufipogon* IRGC 105491 (Orufi). Phenotypes include rachis length (RL; A), exsertion thickness (TE; B), primary branch length (PBL; C), and number of branches per internode (BpN, D). Distributions were calculated using the methods described by Greenberg *et al.* (2011). Briefly, Markov chain Monte Carlo line means were calculated using 1,000 permutations and the panicle phenotypes from plants grown in four different pot sizes (2, 3, 4, and 6 inches). The total number of panicle images was 1,139.

Table 4.5. Comparison of PANorama and P-TRAP software platforms.

	Feature	PANorama	P-TRAP
Software Implementation	Image Acquisition	Built directly into pipeline with skeleton preview	NA
	Skeletonization	Exact morphological erosions of a shape	Axes connected with straight line segments
	Skeletonization	Future software releases can target differently shaped image objects	Generates skeletons for branching structures only
	Axis Definition	Internode/awn filtering built into skeletonization with adjustable parameters	Internodes and awns must be manually edited using a GUI
	Axis Definition	User divides the primary axis three times	User divides primary axis two times
	Axis Definition	Skeleton points are selected on top of a _skel.jpg image, generating an accessory file	Points selected within a skeleton graph and require manual editing of skeleton
	Software Execution	Linear pipeline controlled with scripts	Self contained GUI
	Compatibility	Linux	Windows, Mac OSX, and Linux
Beta Testing	Image number	>2600 images within this study	26 images
	Genetic mapping	QTL Mapping using 154 RILs in *Oryza sativa*	NA
Phenotyping differences	Primary axis internodes	Length of individual rachis internodes	Mean internode length only
	Primary axis thickness	Mean and maximum thickness measures	Mean thickness only
	Branches per internode	Calculated at every internode (mean and maximum)	NA
	Exsertion phenotypes	Length and thickness	NA
	Secondary and tertiary axes	NA	Branch length, count, and number of internodes
	Spikelet traits	NA	Count, length, shape

we demonstrate that PANorama represents an improvement in efficiency, flexibility, and accuracy when phenotyping primary panicle axis traits. Table 4.5 summarizes the major differences in software implementation, testing, and phenotyping capabilities between PANorama and P-TRAP, an open-source software platform released in August 2013 (AL-Tam et al., 2013).

The PANorama pipeline is broken into distinct steps that are executed linearly using a small collection of scripts (Figure 4.2; Figure S1; Videos S1-S3). The most demanding tasks are computed in batch, including awn filtering, skeletonization, and data extraction. Meanwhile, basic orientation of axes and error correction for images are flexibly controlled for each image via standalone GUIs. All parameters controlling implementation of the pipeline are stored in a single editable file (Table S1). PANorama skeletons are calculated using the exact morphology of panicle axes, which accurately excludes awns and requires no editing by the user (Figures 4.2-4.4). When comparing PANorama phenotypes and manual measurements, we found that using the awn filter always increased accuracy of primary branch length (PBL) measurements—even in varieties with developmentally small awns (Table 4.2; Figure S2). For landraces and wild relatives of *Oryza sativa*, which have long awns of varying length (Figure 4.2), this feature was crucial when measuring PBL in high replication (Figure 4.7). When identifying developmental transition points in an image using the PANorama GUI DefineMainAxis.py, points are selected on top of a skeleton and their positions are stored in an accessory .txt file separately from the skeleton (Figure S1; Video S3). This allows a user to flexibly edit any parameters controlling the pipeline (Table SI) and rerun skeleton calculation, without having to re-select developmental transition points for every image in a directory. Selecting three points during this step also

allows division of the main axis into two types of tissue; in this study, vegetative (exsertion length) and reproductive (rachis length).

In contrast, the P-TRAP platform calculates skeletons using straight-line segments and operates entirely within a single GUI, requiring users to manually edit skeletons using a mouse (Table 4.5). P-TRAP GUI implementation is flexible and facilitates large-scale error correction of an image skeleton, but it comes at a great cost in terms of accuracy of skeletons and overall processing time—defeating the purpose of high-throughput image analysis. When comparing PANorama and P-TRAP on the same computer using 60 images, we found that skeleton calculation and phenotype extraction were eight times faster in PANorama (Table 4.2). Additionally, rerunning skeleton calculation within P-TRAP required that the IM initiation and abortion points be re-selected for every image—increasing manual processing time when replicating results (data not included in time calculations). P-TRAP could not distinguish curved panicle axes, cluster branches into internodes automatically, or filter awns (Figure S3). This resulted in decreased accuracy (low r^2 values, large MPEs) and increased error (large MAPDs) for all P-TRAP phenotypes (Table 4.2).

The flexible implementation of skeletonization and awn filtering within PANorama directly facilitated our ability to perform the first large scale, image-based quantification of primary panicle architecture in rice using thousands of images (Table 4.5). In contrast, P-TRAP was only tested on 26 images before being released (AL-Tam et al., 2013). We found our methods to be robust, even when photographing panicles from different varieties and subpopulations. With minimal tweaking to parameters within iftPANorama.h (see Materials and Methods), we were

also able to generate skeletons for plant organs from other species (Figure 4.5; Table S3). For structures like the maize male inflorescence, which is similar to rice, processing only required that the AWNSTHICK parameter be turned off. However, while we could accurately generate skeletons for the Arabidopsis and tomato images (Figure 4.5), we could only extract phenotypes that were similar in structure to a rice panicle (Table S3). By releasing the PANorama source code freely to public, other research groups can expand on the skeleton implementation used within the pipeline and fully customize PANorama to extract measurements from differently shaped image objects.

Quantitative Panicle Phenotypes are Necessary to Leverage High-Resolution Genotype Data

This study represents the first QTL analysis to use both high-density genetic marker data and high-resolution phenotypes to characterize multiple primary panicle architecture in rice. In doing so, we were able to detect the largest number of panicle and exsertion QTL ever reported in a single study (Table 4.3). Several of these QTL have been identified in previous studies using different mapping populations, but many are novel (Table 4.4; Table S4). Using PANorama we were able to measure many traits simultaneously in high-replication, identify new QTL for known traits that have never been identified, and measure several novel phenotypes that are otherwise difficult to capture, such as internode composition of the main rachis axis. Additionally, we identified several regions of the genome that show pleiotropic clustering of panicle QTL, and we confirm overlap of PANorama QTL with two independently measured reproductive traits (DTH and PN). Both of these findings are consistent with previous studies in rice (Kobayashi et al., 2003; Ando et al., 2008; Liu et al., 2008; Ikeda et al., 2010), adding

further biological validity to our QTL analysis. Taken together, these results attest to both the efficiency and accuracy of the PANorama pipeline.

When using high-resolution GBS marker data to map PL (Figure 4.1), we were only able to detect one QTL on chromosome 1. However, by dissecting this composite phenotype into its internode components using PANorama, we detected 16 additional QTL at 11 distinct locations throughout the genome (Figure 4.6B; Table 4.4). Although some of these QTL overlap with one another, several are unique to different rachis measures and together explain larger PVEs (Table 4.3). These findings not only validate our decision to treat the rachis axis as a composite phenotype, but suggest that high-resolution phenotyping is necessary to dissect pleiotropic panicle traits (Figure 4.6B; Table 4.4). Interestingly, the panicle length, rachis length, internode and length phenotypes all have QTL that tightly cluster at the distal end of chromosome 1, in the same region as a QTL for the vegetative trait exsertion length (Figure 4.6B; Table 4.4). A partial measurement of the culm (Figure 4.1), Exsertion length has become an important phenotype with the advent of hybrid rice breeding. Nearly every male-sterile line available to breeders has short EL, which severely reduces outcrossing because panicles remains trapped within the leaf sheath. Even though a gene controlling panicle exsertion has been cloned on chromosome 5, *ELONGATED UPPERMOST INTERNODE* (*EUI*), many breeders still rely on exogenously applied gibberellic acid to ensure full panicle exsertion during crossing schemes (Luo et al., 2006).

The exsertion length LOD profile across chromosome 1 is relatively narrow and highly significant, with one major peak near the gibberellin oxidase gene *SD1* (Figure 4.6C; Table 4.3).

During the Green Revolution, breeders incorporated an allele of *SD1* into diverse *indica* backgrounds that significantly increased harvest index, by reducing the ratio of vegetative to reproductive tissue (Hedden, 2003). This was accomplished by selecting for the phenotype "plant height," measured as the total length of a panicle bearing culm, from soil level at the base of the vegetative stem to the tip of the longest panicle (International Plant Genetic Resources Institute and Association, 2007). This measurement integrates multiple internodes, vegetative and reproductive, into a single composite trait that encompasses all exsertion and panicle phenotypes. As such, it is not surprising that previous QTL studies have mapped PL and plant height to the distal end of chromosome 1, using a population of similar genetic background; IR64 is one of the *indica* varieties that received the dwarfing *SD1* allele, while Azucena is a tall *tropical japonica* landrace. Breeders have always assumed that the *SD1* locus was pleiotropically responsible for length phenotypes that map to this region, even though these studies were conducted using low-density marker coverage and low resolution phenotyping (Hittalmani et al., 2003; Li et al., 2003).

By comparing QTL for panicle length, rachis length, and the various internode measures surrounding *SD1* within our study, we note significant changes with respect to both the overall LOD scores and the distribution of the LOD profiles. Specifically, with increasing resolution of panicle length phenotypes (from panicle length to rachis length to internode length), the LOD profiles resolve from one amorphous peak into several small-QTL peaks (Figure 4.6C). This suggests that that there are likely multiple, linked loci across the region involved in panicle development surrounding the large-effect *SD1* gene, which are easier to detect using a combination of dense marker data and high-resolution panicle phenotype data. Although our

ability to identify specific candidate genes within complex QTL regions is limited by the number of recombination events within this relatively small RIL population, mapping numerous internode-derived traits to this region is consistent with the recent breeding history at this locus. These results present compelling evidence that high-resolution panicle phenotyping will facilitate dissection of complex loci and complement other high-resolution genotype data sets, such as those used in GWAS.

Rice Panicle Architecture Varies Across Different Genetic Backgrounds

The high-resolution phenotypes we generated using PANorama allowed us to detect subtle variation in panicle architecture, which may otherwise have been lost using standard phenotyping methodologies. Many panicle traits have distinctly different distributions within a genetic line, despite being grown in varying pot sizes (Figure 4.7; Table S5-6). These results support findings observed in a previous study, which identified subpopulation-specific polymorphisms associated with panicle length in *O. sativa* using GWAS (Zhao et al., 2011). However, it is interesting to note that although three of the lines in this study (Azucena, Moroberekan, Jefferson) are from the same subpopulation, *tropical japonica*, their panicle traits are distinctly different from one another. This confirms that quantitative variation exists within multiple stages of panicle development—at the variety, subpopulation, and *O. sativa* species level.

Within the QTL mapping portion of this paper, we identify several genomic regions that have overlapping QTL for panicle traits, panicle number, and/or days to heading phenotypes (Table S4). The concept of pleiotropy in panicle development has been confirmed numerous times

within other QTL studies (Kobayashi et al., 2003; Ando et al., 2008; Liu et al., 2008; Ikeda et al., 2010). However, limited genotype and phenotype resolution has always prevented dissection of these complex loci, which could be comprised of single large-effect genes, or arrays of linked, small-effect genes as we suggest exists on chromosome 1 (Figure 4.6C). In either case, without performing high-resolution panicle phenotyping on a larger set of rice germplasm, it is difficult to interpret whether multi-trait loci observed within experimental genetic mapping populations are actually responsible for variety and subpopulation-specific panicle traits, such as those observed within our small collection of genetic lines (Figure 4.7; Table S5). For breeders working with diverse sets of germplasm, the existence of such loci could have significant implications when selecting for panicle-related traits that affect yield and grain quality.

Conclusion

PANorama is a unique and flexible phenotyping pipeline that improved our ability to quantitatively measure rice panicles in both a QTL mapping study and a collection of diverse genetic lines. Although breeding for rice panicle morphology to boost yield potential is promising in theory, quantitative phenotyping must be done on a larger set of germplasm in order to tease apart the pleiotropic nature of panicle phenotypes observed in genetic mapping populations. By releasing PANorama freely to the public, it is our hope that others will be able to adapt it to their own phenotyping needs. The use of a single phenotyping platform across multiple research groups could represent a significant step forward in streamlining comparative genetic studies, both within and across species.

Acknowledgements

We would like to thank the laboratories of Michael Scanlon, Jian Hua, and Joceyln Rose at Cornell University for supplying plant tissue specimens from maize, Arabidopsis, and tomato, respectively. We are grateful to Randy Clark for support in setting up a plant imaging station, to Sandra Harrington and Gen Onishi for support with greenhouse management, and to Joseph LeCates for assistance in harvesting the growth chamber experiment.

Supplemental Materials

The following supplemental data is available in the online version of this article. The open source software PANorama is free to download at ricediversity.org, and videos can be using from the links below.

Supplementary Methods. Image Analysis Methods for High Resolution Inflorescence Phenotyping.

Figure S1. Directory structure used in the PANorama Pipeline.

Figure S2. Final skeletons in closed versus open rice panicles.

Figure S3. Screenshots comparing PANorama to the opensource platform P-TRAP.

Table S1. Parameters stored within the iftPANOrama.h file.

Table S2. Correlation coefficients between PANorama phenotypes with and without awn filtering.

Table S3. Phenotypes from other species.

Table S4. Panicle trait QTL identified in an IR64 x Azucena RIL Population.

Table S5. Phenotypic grand means for 8 genetic lines, calculated using all pot sizes.

Table S6. Phenotypic raw means for each genetic line, calculated per pot size.

Video S1. Using GetPictures.py and RebuildImageFiles.py.

(https://vimeo.com/87497670)

Video S2. Using CreateSkeletons.py and DefineMainAxis.py.

(https://vimeo.com/87497348)

Video S3. Using ExtractInfo.py, CorrectBranches.py, and ExtractMeasures.py.

(https://vimeo.com/87497347)

4.4 Materials and Methods

Plant Material and Harvesting

The set of *Oryza sativa* genetic lines used in the pot size experiment are the parents of 4 mapping populations currently available within the international rice community. Genotypes include Azucena, CO39, IR64, Jefferson, Kasalath, Moroberekan, Nipponbare, and one wild relative (*Oryza rufipogon* IRGC 105491). The IR64 x Azucena RIL population, comprised of 176 F10-F12 lines, was originally developed at IRD (Montpellier, France) using single seed descent (SSD) under greenhouse conditions. For the Comparison of Manual Measures, PANorama, and P-TRAP, IR64 and Azucena varieties from Rice Diversity Panel 1 (RDP1) were used.

All rice seeds used within this study were obtained from S. McCouch. Seeds were surface sterilized using 20% bleach for 15 to 20 minutes, and germinated directly within a soil at ~1cm depth. A 1:1 soil mixture of Promix:Clay was used in both studies. Panicles were harvested by cutting the culm at the flag leaf ligule, so as to include the exsertion length (EL) measurement. For panicles that still remained wrapped within the leaf sheath of the culm, panicles were cut just below the panicle notch (Figure 4.1) so as to include all reproductive phenotypes. Samples were stored in 70% ethanol to maintain flexibility of tissue until time of imaging. Panicle number (PN) was measured as the total number of panicles on a plant after grain filling had finished. Days to heading (DTH) was recorded when the first panicle on a plant emerged 50% from the leaf sheath.

Planting conditions

Growth chamber

Plants were grown in the Weill Growth Chamber Facility at Cornell University, Ithaca, NY. Chamber settings were as follows: 300uM light on a 12 hour day/night cycle; day/night temperature of 30°C/25°C; humidity between 75-85%. The 8 genetic lines were planted in 4 pot sizes (2,3,4, and 6-inch) and replicated (n=40 plants per variety; 10 replicates per pot size). To reduce shading effects, plants were grown in a pseudo-randomized block design. The lines were grouped by plant height (short and tall), and pots were also clustered into two major groups based on size (4 and 6-inch pots together, and the 2 and 3-inch Deepots together). Position of plants within clusters was randomized, and clusters of different sized pots were randomized throughout the chamber. When available, 5 panicles/plant replicate were photographed (total panicles in study: N=1139).

Greenhouse

For "Comparison of Manual Measures, PANorama, and P-TRAP," IR64 and Azucena were grown in summer 2013 in the Gutterman Greenhouse Facility at Cornell University, Ithaca, NY (n=3 plants/variety). Five panicles were harvested from each replicate (N=15 panicles/variety). For the QTL mapping study, the recombinant inbred lines were grown summer 2012 in the Gutterman Greenhouse Facility. In order to reduce shading or greenhouse positional effects, three replicates (n=3 plants/line) were grown in a pseudo-randomized block design. Briefly, plants were grown in 6" pots and clustered by height; plant position within a cluster and position of overall clusters within the greenhouse were both randomized at the start of the study. Several weeks after tillering phase began, plants with excessively short phenotypes or narrow leaves were moved into groups outside of the randomized design to prevent shading. When available, 5 panicles/plant replicate were photographed using PANorama (total panicles in study: N=1537).

Sample Imaging Protocol

Rice

Before photographing, ethanol-soaked panicles were air-dried for 15-24 hours to improve the rigidity of the tissue without resulting in brittleness. Samples were photographed using a Nikon D200 digital SLR camera, a 60mm Macro lens, and a 0.3 neutral density (ND) filter. Camera control in PANorama is mediated by gPhoto2 (version 2.4.14). During imaging, panicles were arranged so that primary branches were spread open so as not to overlap within the image view. The camera was mounted on a fixed copy stand, and color images of samples were acquired on a light box to improve contrast. Images were saved at 2592x3872 pixels. The scale parameter within iftPANorama.h was adjusted for each experiment to ensure that pixel to length conversions were accurate (Comparison of Manual Measures, PANorama, and P-TRAP=116 pixels/cm; QTL mapping=120 pixels/cm; measurement of 8 genetic lines=120 pixels/cm). To determine the scale, an image of a ruler was acquired using GetPictures.py, and the open-source software ImageJ (version 1.46a-1) was used to measure the length of 1cm in pixels at several places within the image. All image processing was performed within the PANorama pipeline, using the parameters in Table S1.

Maize, Tomato, and Arabidopsis

Images for maize, tomato, and Arabidopsis were acquired using PANorama and saved at 2592x3872 pixels (scale=120 pixels/cm). Several adjustments were made to the iftPANorama.h file before skeleton calculation. The segmentation and multi-scale parameters used by CreateSkeletons.py were not altered, but the awn filter was turned off (AWNSTHICK=0.0cm). All other parameters were kept the same for the maize and tomato inflorescences. For the Arabidopsis plant and tomato leaf, the minimum junction distance and minimum primary branch length parameters were altered (Arabidopsis: MINJCTDIST=0.5, MINBRANCSZ=0.1; tomato leaf: MINJCTDIST=0.1, MINBRANCSZ=0.1).

The EP, IMAP, and FMAP points selected in the DefineMainAxis.py GUI were carefully chosen to highlight different structures within each image (Figure 5, color differences). Measurements are stored in Table S3. To measure total Arabidopsis plant architecture, EP was selected at the tip of the longest root, IMAP at the base of the inflorescence, and FMAP at the tip of the inflorescence (Figure 4.5A). To measure Arabidopsis inflorescence traits, the EP was selected at the base of the inflorescence, the IMAP just below the terminal floret cluster, and the FMAP at the tip of the inflorescence (image not shown). For the maize inflorescence, EP was selected at the flag leaf ligule attachment point, IMAP at the first tassel branch, and FMAP at the base of the terminal spike (Figure 4.5B). For the tomato inflorescence, FMAP was selected at the base of the last fruit on the cyme. For the tomato leaf, FMAP was selected at the base of the terminal leaflet. In both tomato images, EP and IMAP were selected directly on top of one another at the base of the axis, in order to exclude pink exsertion measurements (Figure 4.5C-4.5D).

Comparison of Manual Measures, PANorama, and P-TRAP

Manual measurements were taken for the following phenotypes by hand: rachis length (RL), primary branch number (PBN), internode number (NN), and average number of branches per node (BpN). Average Primary branch length (PBL) was calculated by manually measuring every branch within an image using the "Segmented Line Tool" in ImageJ (version 1.46a-1; pixel conversion=116 pixels/cm). Measurements were taken using images that contained openly arranged panicle branches, to ensure that curved primary branches could be measured as accurately as possible. Phenotypes were calculated from raw skeletons in both PANorama and P-TRAP and used to calculate all reported statistics. No images were corrected after skeleton calculation, except to select the meristem transition points. Because P-TRAP had large percent deviations from manual measures for both RL and NN, we reran the analysis two times with P-TRAP freshly installed on two separate computers (Linux and Macintosh). However, doing so did not significantly improve the r^2, MAPDs, or MPEs. Total processing time for PANorama includes the execution of the following scripts for 60 images: CreateSkeletons.py, ExtractInfo.py, and ExtractMeasures.py. Total processing time for P-TRAP includes skeletonization ("Find Structure") and

phenotype calculation ("Collect Results") for 60 images. The amount of time required to select meristem points was not included in either calculation. Additionally, P-TRAP grain quantification ("Find Grains") was not included in time calculation. Processing times were measured on the same computer for both platforms.

QTL analysis

QTL mapping within this study was performed using 154 of the IR64 x Azucena RIL lines and the software package R/QTL (R version 2.15.1; R/qtl package 1.24.9), following the methods described in (Spindel et al., 2013). Briefly, we performed an initial scan for QTL using Haley-Knott Regression and 1,000 permutations to determine the $\log_{10}$ of the odds (LOD) significance threshold. We then re-scanned for additional QTL, conditioning on peaks detected in the initial scan. Finally, forward selection and backward elimination were used to probe the model space and refine QTL locations. For phenotypes that failed the Shapiro-Wilks Test for normality, a natural log transformation of the raw data was performed before QTL mapping was conducted (Table 4.4; Table S4).

Statistical Analysis of Diverse Lines

To construct distributions of line means and pot size effects on each panicle phenotype, we implemented a multivariate Bayesian hierarchical model(Greenberg et al., 2011). The data contained between one and five individual data points per plant, and between 24 and 32 plants (grown in various pot sizes) per genotype. We analyzed all phenotypes (displayed in Figure 4.5) simultaneously. The posterior distributions of line means were constructed using Markov chain Monte Carlo as described in(Greenberg et al., 2011). We ran 500 iterations of burn-in and saved 25000 samples each from four independent chains. Convergence was verified by inspecting trace plots of the chains.

CHAPTER 5

GENOME-WIDE ASSOCIATION AND HIGH-RESOLUTION PHENOTYPING LINK *ORYZA SATIVA* PANICLE TRAITS TO NUMEROUS, TRAIT-SPECIFIC QTL CLUSTERS

Included with permission from:

Samuel Crowell, Pavel Korniliev, Alexandre Falcão, Abdelbagi Ismail, Glenn Gregorio, Jason Mezey, Susan McCouch. (2015). *Nature Communications*, under review.

Abstract. Panicle architecture in rice is a key selection target when breeding for yield potential and grain quality. However, the genetic basis for panicle development has historically been difficult to quantify due to complex branching patterns, highly correlated developmental pathways, and confounding of geographic adaptation with reproductive traits. Here, we collect high-resolution panicle phenotypes using an open-source image-analysis platform, PANorama, and conduct a genome wide association study using 700,000 single nucleotide polymorphisms. We detect numerous subpopulation-specific associations and clusters of panicle trait QTLs, but note panicle length and count traits generally map independently. Seven candidate genes were identified in hormone pathways known to regulate rice plant architecture. This is the first, high-resolution dissection of inflorescence architecture using field grown material for any species. Herein, we establish a panicle size morphocline for domesticated rice, propose a genetic model underlying complex panicle traits, and demonstrate subtle links between panicle size and yield performance.

5.1 Introduction

As the bearers of grain, grass inflorescences have been the target of selection for thousands of years (Doust, 2007). In Asian rice (*Oryza sativa*), an important staple crop for millions of people,

optimizing rice panicle size and shape represents an important target of selection for breeders attempting to improve yield potential and maximize grain quality (Doust, 2007; Wang and Li, 2011). Overall panicle size and branching patterns in rice have increased in complexity throughout domestication and modern breeding, but changes have been relatively subtle. Seeds are born on long primary panicle branches that sometimes iterate into secondary and tertiary branching structures (Ikeda et al., 2004; Wang and Li, 2011), and these characteristics are often variety-specific and variable under different environmental conditions (Kobayashi et al., 2003; Li et al., 2003; Ikeda et al., 2004). Thus, unlike in maize (*Zea mays*), where inflorescences have been selected for extreme divergence into a branchless female cob and a highly branched male tassel (Kellogg, 2007), the panicles from many modern rice varieties still resemble those from their closest wild relatives, *Oryza rufipogon* and *Oryza nivara* (Yamaki et al., 2010).

Many genes have been cloned relating to inflorescence development in rice (Yoshida and Nagato, 2011; Zhang and Yuan, 2014), several of which are agronomically important. The *OsLIGULESS1* (*OsLG1*) locus was recently identified as a rice domestication gene and controls the shift from open to closed panicles (Ishii et al., 2013). A natural allele of the *DENSE AND ERECT PANICLE 1* (*DEP1*) gene within high-yielding Chinese rice varieties boosts yield potential by pleiotropically reducing panicle internode length, while increasing both primary and secondary branch number (Huang et al., 2009). Additionally, a major effect allele for the *Grain Number 1a* (*GN1a*) gene significantly increases secondary panicle branching, grain count, and yield (Ashikari, 2005) and is already being incorporated into rice breeding pipelines. However, while many studies have examined the role of candidate genes in the reference sequenced variety

(Nipponbare) or its close relatives, panicle architecture has not been characterized in detail across diverse varieties that are grown by farmers.

The inbreeding nature of rice and multiple origins of domestication have led to the formation of deep subpopulation structure, which has partitioned genetic and phenotypic variation. *Oryza sativa* is comprised of two varietal groups (subspecies), *Indica* and *Japonica*, which can be further divided into five subpopulations (*indica, aus, tropical japonica, temperate japonica, aromatic/Group V*) (Garris et al., 2005; Zhao et al., 2010). Several genome-wide association studies (GWAS) have confirmed that variation exists both within and between rice subpopulations for important agronomic traits (Huang et al., 2010; Zhao et al., 2010; Huang et al., 2011), including panicle count and panicle length (Huang et al., 2011; Zhao et al., 2011). However, due to architectural complexity and limitations in phenotyping, rice primary panicle architecture has never been dissected in detail across a wide range of varieties (Crowell et al., 2014).

In this study, we performed GWAS using a genotypic data set consisting of 700,000 single nucleotide polymorphisms (SNPs) assayed using a high density rice array (HDRA) (see companion paper by McCouch et al.) and phenotypes collected with a high-resolution panicle phenotyping platform, PANorama (Crowell et al., 2014). Unlike previous studies, which focused on collecting a few trait measurements in a large population of varieties (Huang et al., 2010; Huang et al., 2011), we collected 60 panicle and agronomic phenotypes on a targeted population of 242 diverse rice varieties grown under field conditions in the tropics. Using this approach, we were able to link several candidate genes to rice panicle development and identify new

quantitative trait loci (QTL) associated with panicle size. Additionally, we demonstrate that panicle traits are governed by a large number of loci that share subtle relationships with other important agronomic traits, phenotypically and genotypically.

5.2 Results

Diversity panel selection and population structure

The phenotyping panel used in this study contained 242 inbred lines representing germplasm from 60 countries, most of which are tropically or subtropically adapted accessions (Supplementary Table 1). Using the Bayesian clustering program *fastStructure* (Raj et al., 2014) we calculated varying levels of K means (Supplementary Figure 1a). The *Indica* and *Japonica* varietal groups appear clearly at K=2, and at K=3 *Indica* further divides into the *indica* and *aus* subpopulations. We found that K=8 best explains the genetic structure within our panel (see Methods), defining clear variation within and between *indica*, *aus, tropical japonica, temperate japonica,* and admixed varieties. Using principle component (PC) analysis we confirmed that the top three PCs account for the *aus, indica,* and *tropical japonica* subpopulations and explain ~30% of the genetic variance within our panel (Supplementary Figure 1b). These results are consistent with previous studies quantifying genetic variation within individual *O. sativa* subpopulations and confirm that our panel captures a range of variation in tropical rice germplasm (Zhao et al., 2010; Huang et al., 2011; Huang et al., 2013).

High-resolution phenotyping reveals relationships between rice panicle traits

Using the image skeletonization phenotyping platform PANorama(Crowell et al., 2014) we measured 49 phenotypes from over 3,400 images of rice panicles collected in the field (Figure

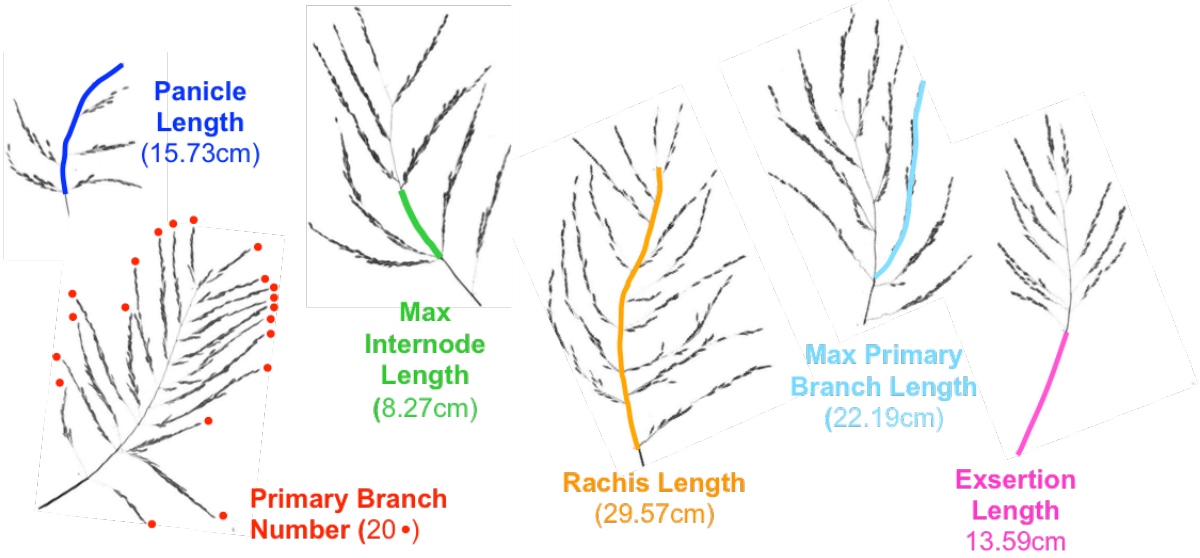
5.1a). The width, length, and count phenotypes extracted from images are nested measurements of three major panicle traits: primary branches, rachis internodes, and culm exsertion (Figure 5.1b). The novel nested measurements incorporated into PANorama in this study are available an updated version of the open source software (Methods). For some of these phenotypes, narrow sense (h^2) and broad sense (H) heritabilities were nearly equivalent, demonstrating the power of image analysis in reducing environmental noise (Supplementary Figure 2). To complement this data set, we collected numerous vegetative and reproductive stage phenotypes, including heading date (a measurement of flowering time). Detailed descriptions of each phenotype are presented in Supplementary Table 2.

Many panicle traits showed highly significant, positive correlations with one another; the median Pearson's correlation coefficient from all pairwise comparisons of panicle phenotypes was r=0.4. Increased rachis and culm width were positively correlated with increases in branch length and branch count traits. Internode number and branch count traits, which are estimates of meristematic divisions, were positively correlated (Figure 5.2a). Groups of similar phenotypes, such as different measurements quantifying branch length traits, were tightly correlated with one another (Supplementary Figure 3). In short, larger panicles always showed thicker axes, longer branches, and higher counts of branches and internodes.

The only strongly negative correlations we detected were those involving internode length traits. Minimum internode length and internode number had an r = -0.81, while minimum internode length and primary branch number had an r = -0.79 (Figure 5.2a). Inverse correlations between length and count traits have been well documented in rice, especially between panicle number

Figure 5.1. Panicle phenotyping in *Oryza sativa*. (a) A diverse collection of landraces were assessed for panicle and agronomic traits under field conditions. A range of extreme panicle phenotypes. depicted in relative scale, highlight a range of phenotypic diversity. (b) The PANorama phenotyping platform generates skeletons from panicle images using exact morphological erosions of shapes. The schematic depicts the major classes of phenotypes extracted from panicle skeletons.

5.1a


Panicle Length (15.73cm)
Max Internode Length (8.27cm)
Primary Branch Number (20 •)
Rachis Length (29.57cm)
Max Primary Branch Length (22.19cm)
Exsertion Length 13.59cm

5.1b

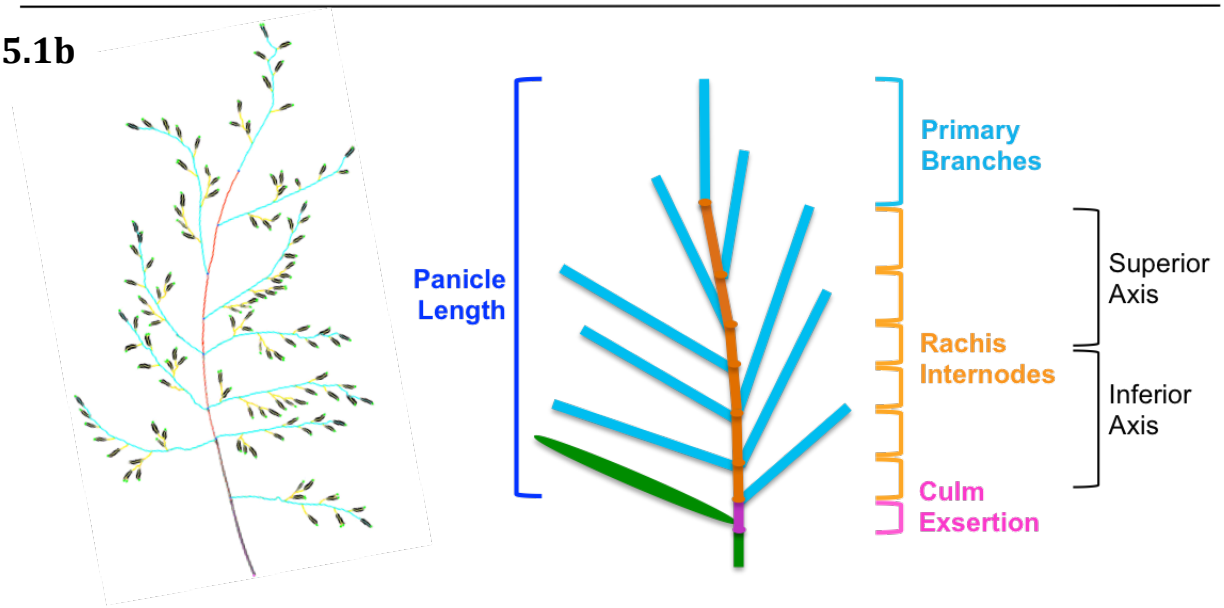
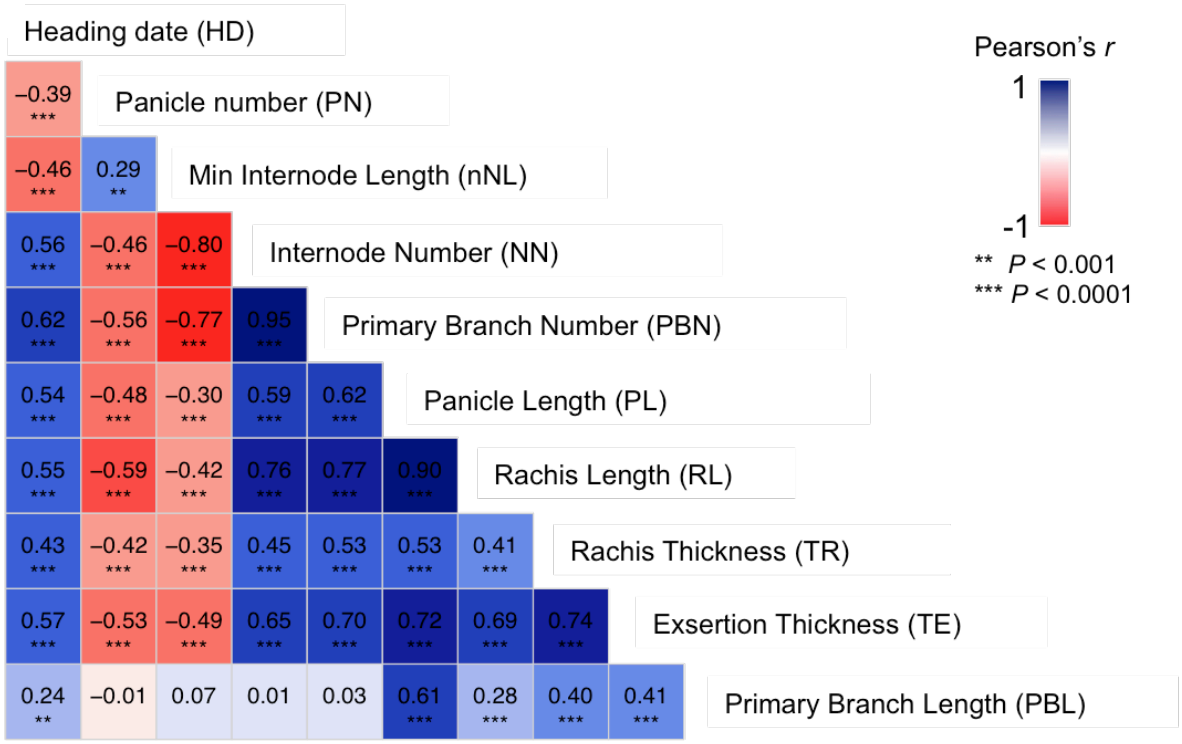
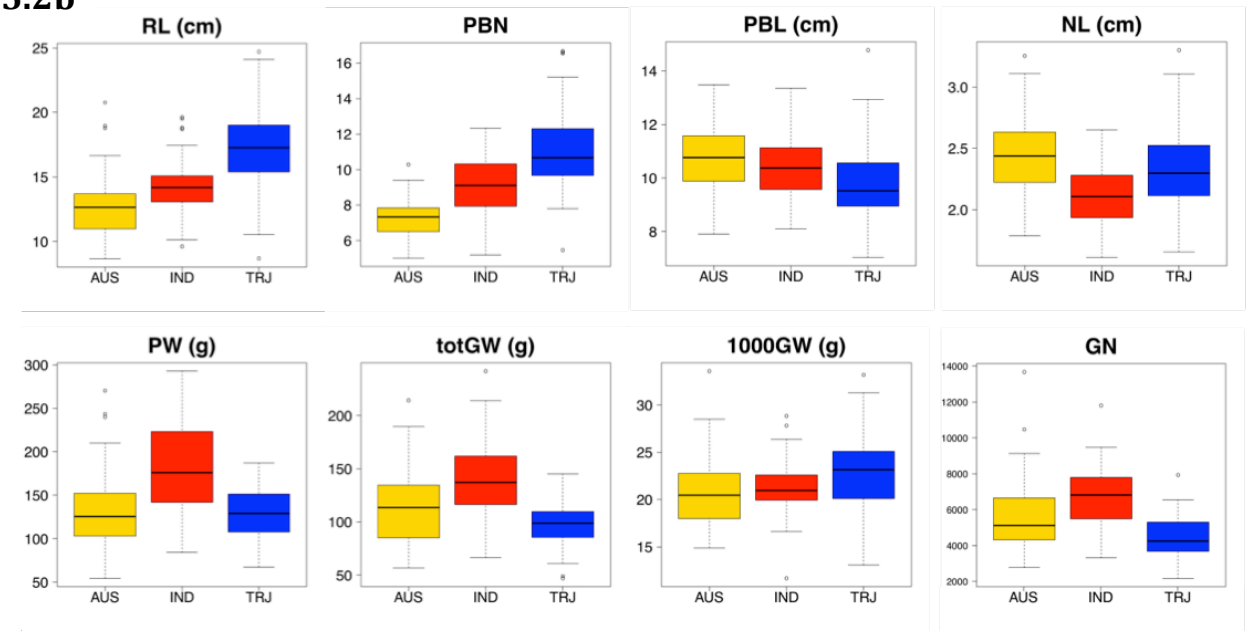

Panicle Length
Primary Branches
Rachis Internodes
Culm Exsertion
Superior Axis
Inferior Axis

Figure 5.2. Phenotypic analysis reveals trait relationships and subpopulation characteristics. (a) A heatmap depicting Pearson's correlation coefficients between a subset of traits within the study. Trait acronyms are in parentheses. A heatmap containing all pairwise comparisons between traits is available in Supplementary Figure 4. (b) Phenotypic distributions of panicle and agronomic traits, divided by subpopulations: rachis length (RL), primary branch number (PBN), primary branch length (PBL), internode length (NL), panicle weight (PW), total grain weight (totGW), 1000-grain weight (1000GW), and grain number (GN). Within boxplots, the bold line represents the median, box edges represent upper and lower quantiles, and whiskers are 1.5 times the quantile of the data. Outliers are shown as open dots.

5.2a


Heading date (HD)
Panicle number (PN)
Min Internode Length (nNL)
Internode Number (NN)
Primary Branch Number (PBN)
Panicle Length (PL)
Rachis Length (RL)
Rachis Thickness (TR)
Exsertion Thickness (TE)
Primary Branch Length (PBL)
Pearson's r
1
-1
** P < 0.001
*** P < 0.0001

5.2b


RL (cm)
PBN
PBL (cm)
NL (cm)
PW (g)
totGW (g)
1000GW (g)
GN
AUS
IND
TRJ

and panicle size (Figure 5.2; Supplementary Figure 3), highlighting that there are physiological and physical tradeoffs within plant development (Ohsumi et al., 2011). As such, we were surprised that we did not observe a tradeoff between panicle branch length and branch count phenotypes (Figure 5.2a). Rather, correlations between these groups of traits were either insignificant or positive (Supplementary Figure 3).

Within individual subpopulations, many panicle phenotypes showed distinct distributions (Figure 5.2b). The *tropical japonica* subpopulation had the highest average panicle length (25cm) and number of branches (11), while the *aus* subpopulation had the largest average branch length (11cm). Historically, many of the highest-yielding varieties have been bred within the *indica* subpopulation (Hedden, 2003); thus, it is not surprising that *indica* outperformed both *aus* and *tropical japonica* in several metrics of yield. However, *indica* varieties generally had intermediate sized panicles and the smallest average internode length. Despite varying distributions among phenotypes within the subpopulations, all correlative relationships between panicle traits and yield components were largely the same in both the *Indica* and *Japonica* varietal groups (Supplementary Figure 3). The highest yielding varieties in our panel never had extreme panicle phenotypes.

Controlling for confounding effects in GWAS: Subpopulation structure and flowering

The inbreeding nature of rice has resulted in the formation of deep subpopulation structure and considerable linkage disequilibrium (LD), which confounds association studies by reducing mapping resolution and increasing type I error (Atwell et al., 2010; Huang et al., 2010; Zhao et al., 2010). Within our panel, average LD does not decay below an r^2=0.2 until ~100 kb in *indica,*

150 kb in *aus,* and 400 kb in *tropical japonica* (Supplementary Figure 4). Additionally, previous GWAS in rice and Arabidopsis have noted that reproductive phenotypes are particularly susceptible to confounding, due to the correlations between flowering time, ecological adaptation, and geographic origin (Atwell et al., 2010; Huang et al., 2011; Zhao et al., 2011). To address these issues, we used a mixed model to correct for subpopulation structure (Kang et al., 2010), integrated the first three PCs as covariates within the model, and performed GWAS across all varieties and within subpopulations individually. Additionally, we repeated all analyses with and without heading date as a covariate. Detailed association results for every trait, subpopulation, and covariate combination are located within the supplemental materials (Supplementary Figures 5-64), as well as at www.ricediversity.org.

The strong, unidirectional correlations observed between many of the panicle phenotypes suggested that they are not genetically or physiologically independent of one another, and that many aspects of panicle morphology are pleiotropic, governed by closely linked genes, and/or co-regulated by a master developmental switch. To explore the underlying reason for the observed correlations, we first examined the role of flowering time on panicle architecture. We hypothesized that close linkage or pleiotropy among genes governing the floral transition and panicle development would manifest in two ways: overlapping associations between panicle traits and heading date, and/or reduced significance of panicle trait associations when controlling for heading date as a covariate in the GWAS model.

We identified five loci associated with heading date across the panel, all of which were detected at low significance (Figure 5.3a). Only one of these peaks, a region on chromosome 2 within the *Indica* varietal group, overlapped with associations for the panicle traits minimum internode length, panicle length, and maximum exsertion thickness (Figure 5.3b). When controlling for heading date as a covariate, the number of significant SNPs associated with panicle traits decreased throughout the genome. Additionally, the peaks for panicle traits on chromosome 2 were attenuated or eliminated (Supplementary Figures 8, 11, and 52), suggesting that the phenology associated with this locus is largely explained by variation in flowering time. Table 5.1 provides a summary of the number of significant associations lost within each subpopulation when using the heading date covariate. The effect was most striking within the *tropical japonica* subpopulation, though several of the *tropical japonica* varieties in our panel share ancestry with *temperate japonica* lines and could be less adapted for growth in the tropics (Supplementary Figure 5.1a). Many of the significant SNPs eliminated when using the heading date covariate were from two peaks within the pericentromeric regions of chromosome 8 (~45 SNPs in *tropical japonica*, 75 when mapping with all varieties) Additionally, many were from primary branch number traits (~130 SNPs), which generally showed improved quantile-quantile plots (Supplementary Figures 33-47).

Although including the heading date covariate in the mixed model eliminated spurious SNPs for several phenotypes, many of the significant SNPs we detected for panicle traits were unaffected (Table 5.1). Most of the peaks associated with length and width traits were not eliminated and occasionally showed increases in significance (Supplementary Figures 9d, 29d, 49d, 50d). Having properly controlled for the effects of flowering time, we further investigated the

Figure 5.3. Genome-wide association results for heading date. (a) Manhattan plots and quantile-quantile plots depicting GWAS results using a mixed model. Associations identified in all varieties, the *Indica* subspecies, and the *Japonica* subspecies are depicted in separate panels. The x-axis depicts the physical location of SNPs across the 12 chromosomes of rice, and the y-axis depicts the $-\log_{10}(P \text{ value})$. Significant SNPs with $P < 1 \times 10^{-5}$ are depicted as colored dots, labeled to match the group in which they were identified (red for *Indica*, blue for *Japonica*). (b) Manhattan plots for chromosome 2. The significant SNP associations for heading date in the *Indica* subspecies are in LD with significant associations for several panicle traits, depicted in separate panels: minimum internode length in the *aus* subpopulation (yellow SNPs); panicle length and maximum exsertion thickness across all varieties (grey SNPs); and maximum exsertion thickness in the *indica* subpopulation (*ind*, red SNPs). (c) An association network summarizing all Manhattan plots in (a) and (b). Traits are labeled with acronyms corresponding to (b). Linkage disequilibrium (LD) blocks are labeled with chromosome number and coordinates. Traits and LD blocks containing significant SNPs are treated as nodes, and are connected if an LD block contains a significant association for the trait of interest. The color and style of the edges connecting the trait and associations indicate which subpopulation or subspecies in which the association was detected. When multiple edges are present between a trait and LD block, a significant association was detected in more than one GWAS. The green arrow indicates the significant peak on chromosome 2 (b), which contains overlapping associations for different types of traits.

5.3a

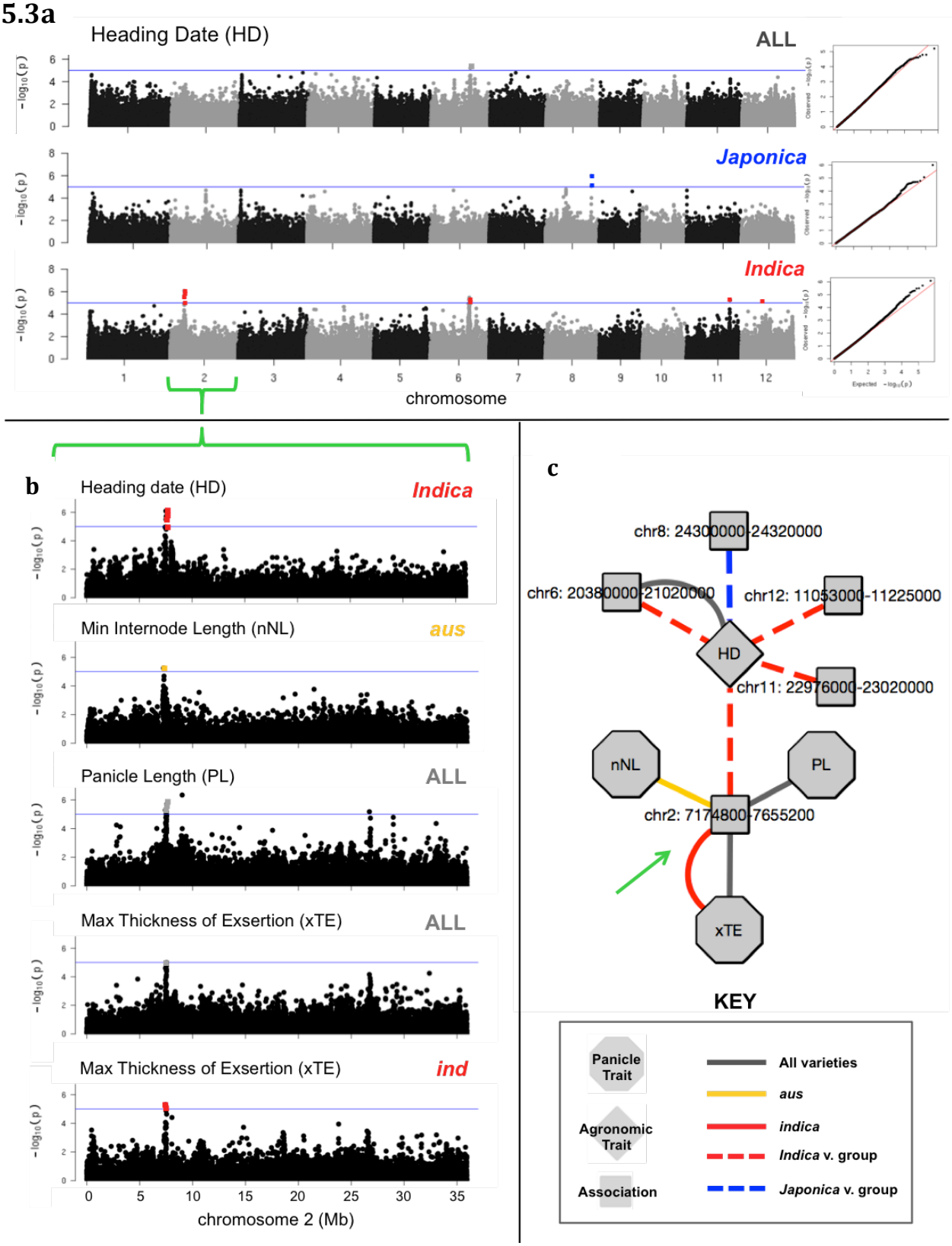

Heading Date (HD)
ALL
Japonica
Indica
chromosome
b
Heading date (HD)
Indica
Min Internode Length (nNL)
aus
Panicle Length (PL)
ALL
Max Thickness of Exsertion (xTE)
ALL
Max Thickness of Exsertion (xTE)
ind
chromosome 2 (Mb)
c
chr8: 24300000-24320000
chr6: 20380000-21020000
chr12: 11053000-11225000
HD
chr11: 22976000-23020000
nNL
PL
chr2: 7174800-7655200
xTE
KEY
Panicle Trait
Agronomic Trait
Association
All varieties
aus
indica
Indica v. group
Japonica v. group

Table 5.1. Genome-wide association results divided by subpopulation and covariate combinations. The number of significant peaks, defined by binning significant SNPs using the a sliding window of linkage disequilbrium (LD), is listed within parentheses.

Subpopulation	Mixed model	Mixed Model + Heading Date Cov.
All	709 (358)	496 (256)
aus	148 (52)	117 (44)
indica	48 (38)	54 (39)
tropical japonica	132 (41)	49 (24)
Total associations	1037 (489)	716 (363)

correlative relationships observed among panicle phenotypes. Thus, unless otherwise noted, all results discussed within the following sections were generated using the heading date covariate.

Association networks facilitate visualization of complex trait relationships

In order to compare association results across many traits, we constructed "association networks" using the program Cytoscape (Shannon et al., 2003). Briefly, significant SNPs were binned based on physical position into peaks, using a sliding window defined by association significance level and local LD (see Methods). We then constructed networks in which traits and peaks were treated as nodes, connected by an edge only when the trait showed significant associations within a given region of the genome. As demonstrated in Figure 5.3c, of the five significant peaks for heading date detected in the genome (Figure 5.3a), only the peak on chromosome 2 shared overlap with panicle traits (as discussed above). Association networks provided a useful visual summary of how peaks were distributed across different traits, and allowed us to quantitatively identify regions of the genome that were significantly associated with different components of phenotypic variation. An interactive file containing all networks discussed within the publication is available at ricediversity.org (Supplementary Data 1).

GWAS links natural variation in panicle traits with numerous, low significance loci

When mapping across all accessions within the panel using the heading covariate, we detected 496 significant SNP associations clustered under 256 peaks located on all 12 chromosomes (Table 5.1). Many SNPs had small to intermediate significance levels; only 18 SNPs showed a $P < 1 \times 10^{-7}$, and the most significant panicle trait association was for primary branch length

standard deviation (P = 8.2 x 10^{-9}; Supplementary Figure 26). These results suggest that many genes could govern panicle morphology, each with small effect.

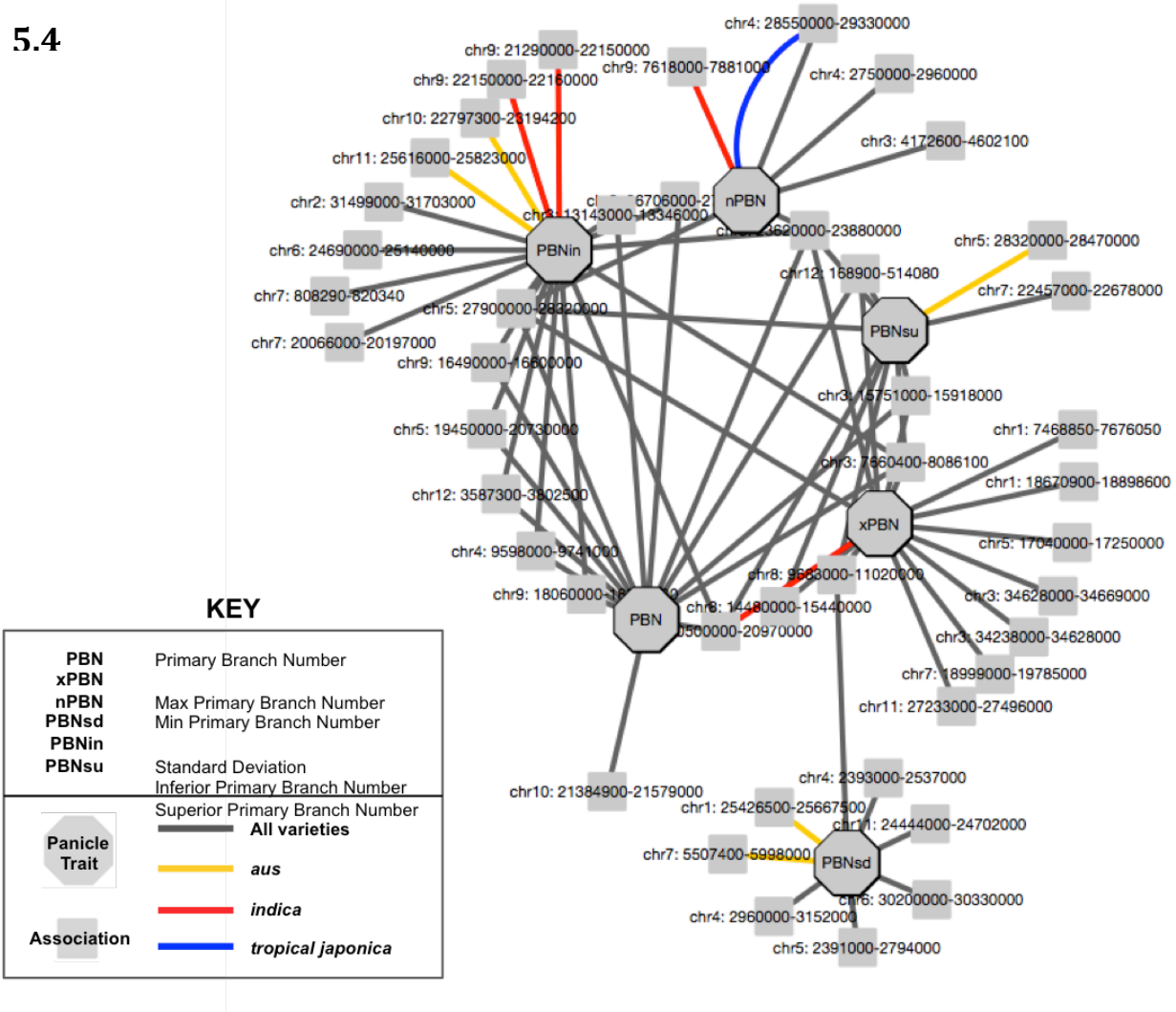
While nested phenotypes often shared the same significant SNPs, we also detected an increased number of significant peaks by dividing panicles into numerous sub-traits. For example, we identified 14 significant peaks when mapping for average number of primary branches across all accessions (Figure 5.4). Mapping with maximum, minimum, and standard deviation of primary branch number identified an additional 15 peaks on 7 chromosomes that were not detected when mapping with average primary branch number (Figure 5.4; Supplementary Figures 33-38). Previous research has shown that there is differential regulation and expression of proteins between spikelets located on lower versus upper panicle branches (Zhang et al., 2014). By mapping for primary branch number in the inferior and superior halves of the panicle separately (Figure 5.1b), we identified an additional 5 peaks not observed among other primary branch traits. We achieved similar results when partitioning other traits into multiple phenotypes (Supplementary Figures 16-28, 39-47; Supplementary Data 1). Thus, although clusters of related measurements are highly correlated with one another morphologically and genetically (Figures 5.2 and 5.4), separating a trait into nested phenotypes resolves the location of small-effect QTL in unique regions of the genome (Crowell et al., 2014).

Subpopulation specific heterogeneity

Performing GWAS within individual subpopulations identified an additional 107 significant peaks, highlighting that genetic variation has either been partitioned or arisen *de novo* during domestication and geographic radiation of *O. sativa*. When comparing significant peaks using

Figure 5.4. Genome-wide association links numerous loci to variation in panicle traits. An association network constructed using primary branch number traits. Traits and linkage disequilibrium (LD) blocks containing significant SNPs ($P < 1 \times 10^{-5}$) are treated as nodes, and are connected if an LD block contains a significant association for the trait of interest. LD blocks are labeled with chromosome number and coordinates. The color and style of the edges connecting the trait and associations indicate which subpopulation or subspecies in which the association was detected. When multiple edges are present between a trait and LD block, significant associations were detected in multiple populations.

5.4


KEY
PBN
xPBN
nPBN
PBNsd
PBNin
PBNsu
Primary Branch Number
Max Primary Branch Number
Min Primary Branch Number
Standard Deviation
Inferior Primary Branch Number
Superior Primary Branch Number
Panicle Trait
Association
All varieties
aus
indica
tropical japonica
chr4: 28550000-29330000
chr9: 21290000-22150000
chr9: 22150000-22160000
chr9: 7618000-7881000
chr4: 2750000-2960000
chr10: 22797300-23194200
chr3: 4172600-4602100
chr11: 25616000-25823000
chr2: 31499000-31703000
nPBN
PBNin
chr6: 24690000-25140000
chr5: 28320000-28470000
chr12: 168900-514080
chr7: 808290-820340
chr7: 22457000-22678000
chr5: 27900000-28320000
PBNsu
chr7: 20066000-20197000
chr9: 16490000-16600000
chr3: 15751000-15918000
chr5: 19450000-20730000
chr1: 7468850-7676050
chr3: 7660400-8086100
chr1: 18670900-18898600
chr12: 3587300-3802500
xPBN
chr5: 17040000-17250000
chr4: 9598000-9741000
chr3: 34628000-34669000
chr8: 14480000-15440000
PBN
chr3: 34238000-34628000
chr7: 18999000-19785000
chr11: 27233000-27496000
chr10: 21384900-21579000
chr4: 2393000-2537000
chr1: 25426500-25667500
chr11: 24444000-24702000
chr7: 5507400-5998000
PBNsd
chr6: 30200000-30330000
chr4: 2960000-3152000
chr5: 2391000-2794000

association networks, we noted that certain types of traits showed enrichment for subpopulation-specific SNPs. For example, while we only identified 10 subpopulation-specific peaks for primary branch count traits (Figure 5.4), we identified 23 peaks for primary branch length traits (Supplementary Figure 65). Strikingly, no two subpopulations had a significant peak for the same trait within the same region of the genome (Supplementary Data 1). Both of these observations have previously been made within rice for many phenotypes, including panicle length (Huang et al., 2010; Zhao et al., 2010; Huang et al., 2011; Zhao et al., 2011), and demonstrate that genetic heterogeneity within *O. sativa* drives trait variation at the subpopulation and subspecies level. Recombination of this variation using wide crosses between varieties from different subpopulations or subspecies provides a mechanism for driving transgressive phenotypic variation. As such, identification of loci that simultaneously impact multiple traits represents a strategy for targeting selection in rice yield improvement pipelines.

Evidence for pleiotropy or genetic linkages between traits

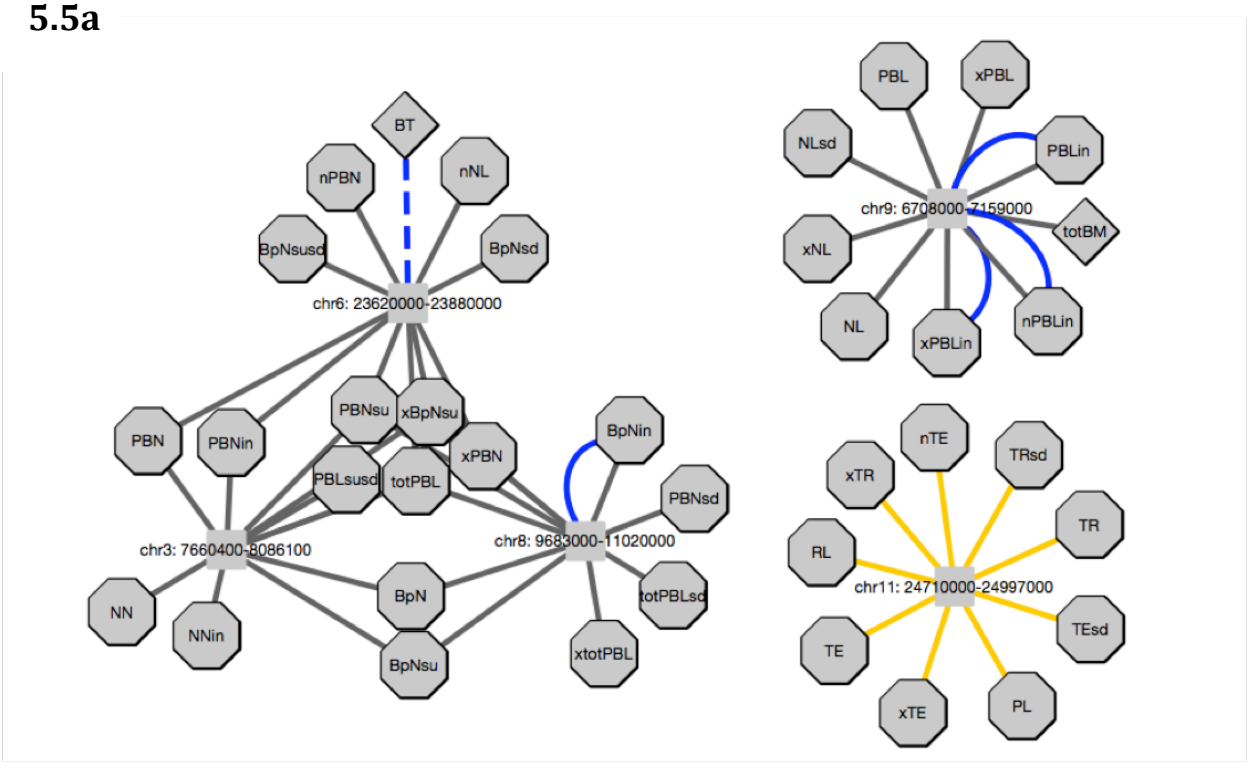
In order to determine whether correlations between different types of traits (length, width count) were the result of significant SNPs mapping to the same LD block, we assessed association networks constructed using every phenotype in the panel. We observed 92 regions of the genome with significant SNPs for more than one trait; 10 of these regions had significant associations for 8 or more panicle traits (Supplementary Data 1). In general, we noted that the same types of traits were most likely to have overlapping peaks within a given region of the genome. For example, two regions on chromosomes 9 and 11 had associations for several internode and branch length traits, which were distinct from peaks on chromosomes 3 and 8 containing associations for count traits (Figure 5.5a). This suggested that panicle traits with a shared

morphological origin, expansion of tissue versus division of meristems, are more likely to be co-inherited due to either pleiotropy or genetic linkage.

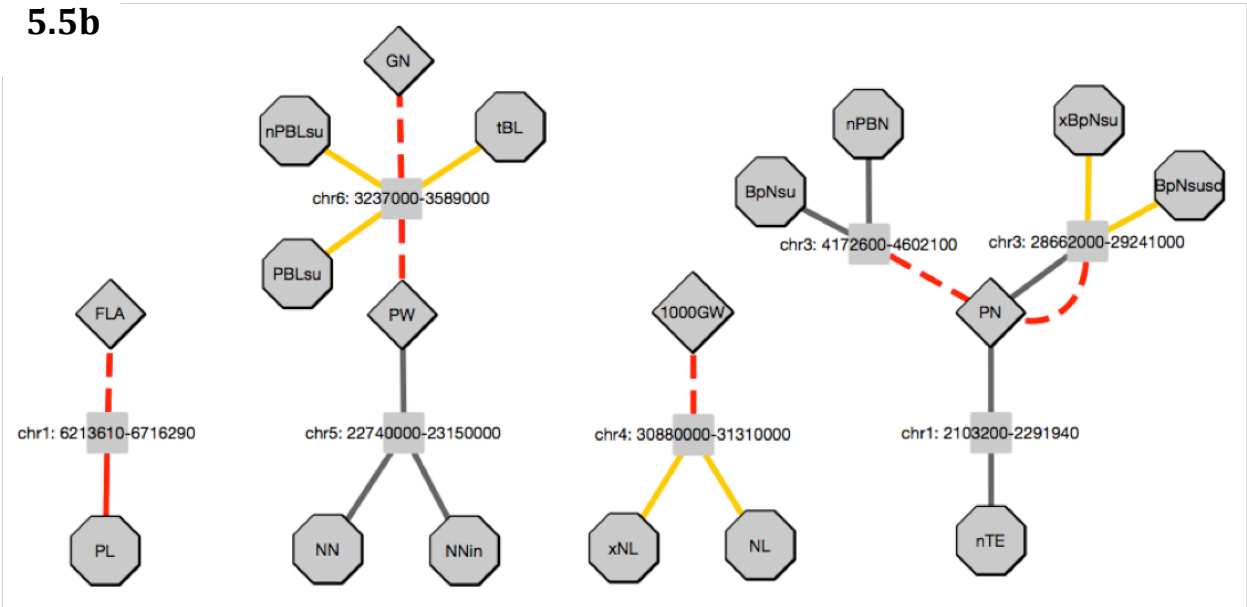
The majority of genomic regions associated with more than one trait were only detectable when mapping across all varieties in the panel, though we occasionally detected subpopulation-specific associations for multiple traits at one genomic address. For example, a QTL on chromosome 11 within the *aus* subpopulation contained associations for several traits with relative effect sizes that were above 30% (Figure 5.5a; Supplementary Table 3). However, overlap of associations from different subpopulations was rare regardless of the traits being compared, and we never observed associations for the same types of traits in an LD block in different subpopulations. When controlling for heading date as a covariate, we detected only one region of the genome that was significantly associated with panicle traits identified in two subpopulations (Figure 5.6a). An LD block located at the distal end of chromosome 10 contained significant associations for a primary branch number trait in *aus* and several primary branch length traits *tropical japonica* (Supplementary Figures 20, 22, 34). While these peaks were distinct from one another (Figure 5.6a), their most significant SNPs were only ~126 kb apart and well within the average range of LD in our panel (Supplementary Figure 4). Taken together, these results suggest that a sizable portion of the genetic variation responsible for panicle morphology remains isolated within respective subpopulations. Attempts to introgress QTL between varieties from different subpopulations using wide crosses would inevitably impact traits in a background-specific manner (Famoso et al., 2011; Zhao et al., 2011), especially if the footprint of a QTL were large enough to encompass multiple alleles that affect the trait of interest (Crowell et al., 2014).

Figure 5.5. Summary of genomic regions containing a large number of overlapping associations for different types of traits. (a) Association networks demonstrate that clusters of related traits frequently have overlapping associations within linkage disequilibrium (LD) blocks. Traits and LD blocks containing significant SNPs ($P < 1 \times 10^{-5}$) are treated as nodes, and are connected if an LD block contains a significant association for the trait of interest. LD blocks are labeled with chromosome number and coordinates. The color and style of the edges connecting the trait and associations indicate the subpopulation or subspecies in which the association was detected. When multiple edges are present between a trait and LD block, a significant association was detected in more than one GWAS. To simplify networks, acronyms are used. Agronomic traits: booting (BT) and total shoot biomass (totBM). Count traits: primary branch number (PBN), internode number (NN), and number of branches per internode (BpN). Length traits: panicle length (PL), rachis length (RL), primary branch length (PBL), and internode length (NL). Width traits: thickness of rachis (TR) and thickness of exsertion (TE). In addition to overall trait averages, sub-traits are depicted using acronyms with prefixes and/or suffixes: maximum (x-), minimum (n-), standard deviation (-sd), inferior (-in), and superior (-su). For example, maximum primary branch length in the inferior half of the panicle (xPBLin). (**b**) Association networks show overlaps between agronomic traits and panicle traits. Agronomic traits: flag leaf area (FLA), panicle weight (PW), grain number (GN), 1000-grain weight (1000GW), and panicle number (PN).

5.5a


BT
nPBN
nNL
BpNsusd
BpNsd
chr6: 23620000-23880000
PBNsu
xBpNsu
PBN
PBNin
xPBN
PBLsusd
totPBL
BpNin
PBNsd
chr3: 7660400-8086100
chr8: 9683000-11020000
NN
NNin
BpN
totPBLsd
BpNsu
xtotPBL
PBL
xPBL
NLsd
PBLin
chr9: 6708000-7159000
xNL
totBM
NL
xPBLin
nPBLin
nTE
TRsd
xTR
TR
RL
chr11: 24710000-24997000
TE
TEsd
xTE
PL

5.5b


GN
nPBLsu
tBL
chr6: 3237000-3589000
PBLsu
PW
FLA
chr1: 6213610-6716290
chr5: 22740000-23150000
PL
NN
NNin
1000GW
chr4: 30880000-31310000
xNL
NL
nPBN
xBpNsu
BpNsusd
BpNsu
chr3: 4172600-4602100
chr3: 28662000-29241000
PN
chr1: 2103200-2291940
nTE

KEY

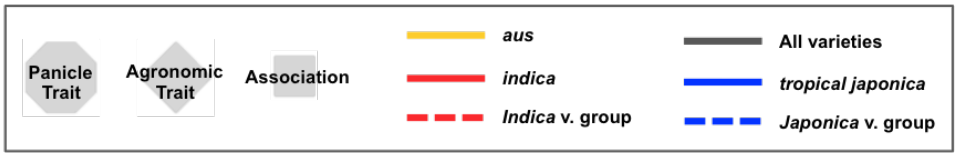
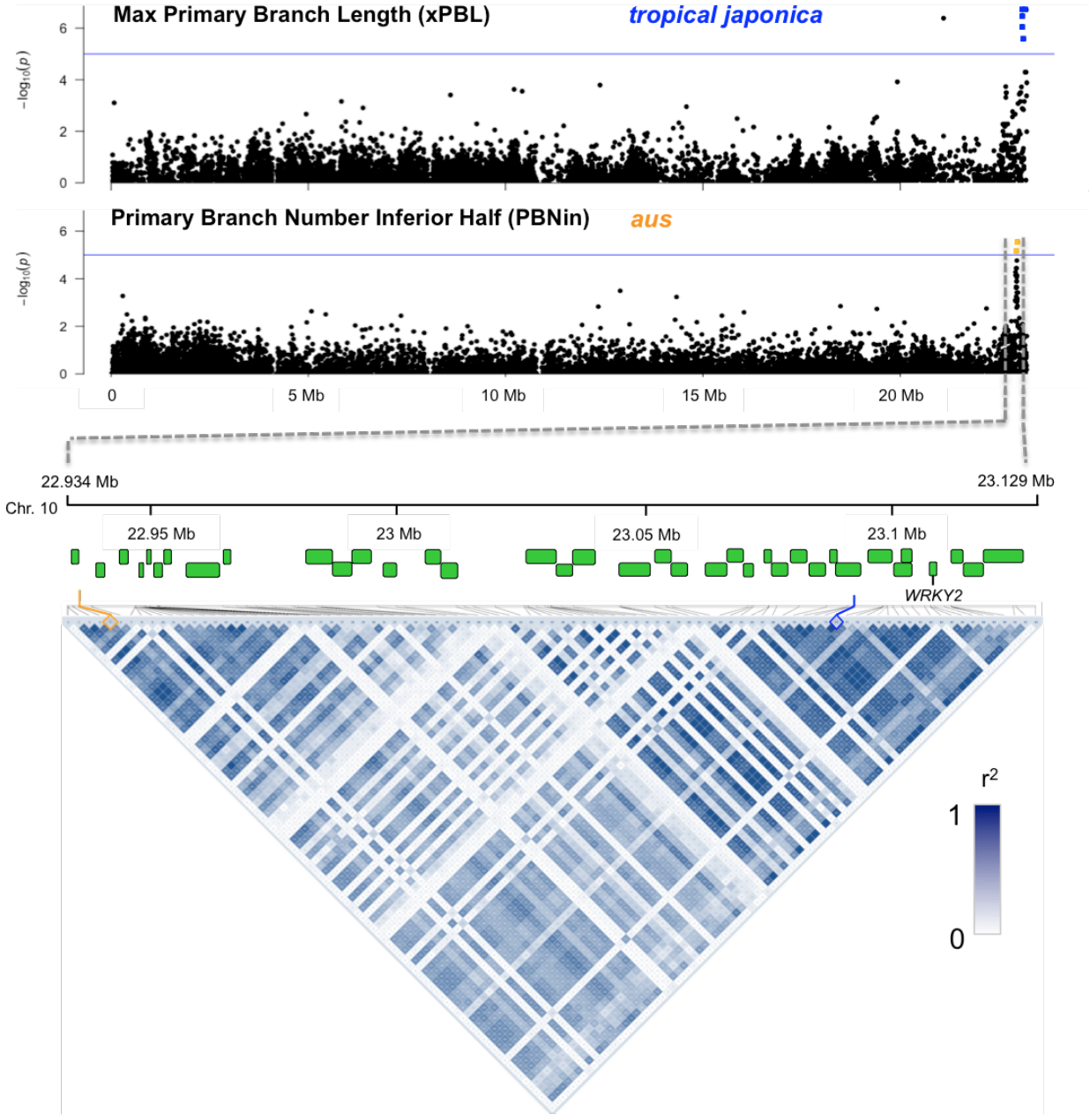
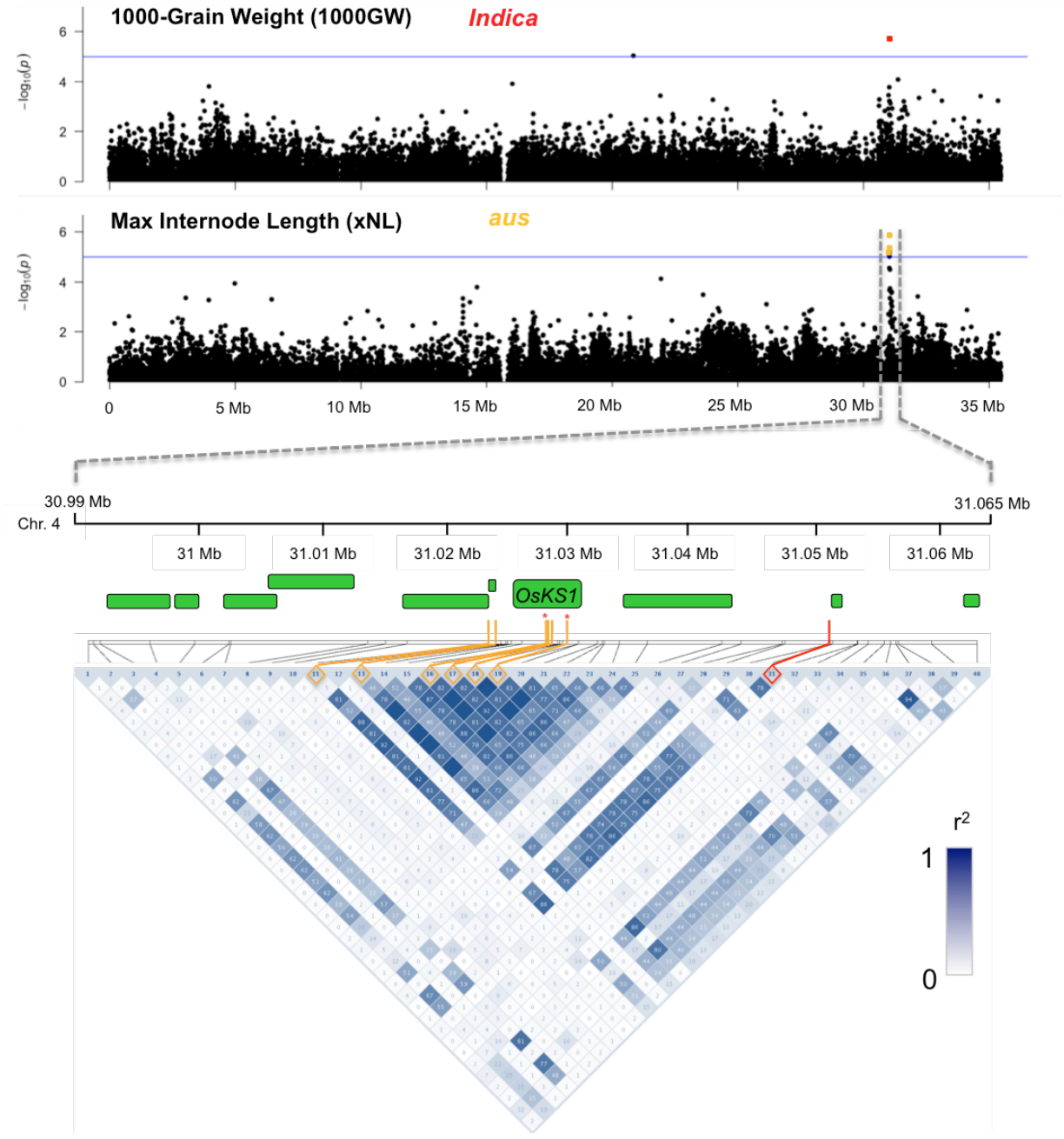

Panicle Trait
Agronomic Trait
Association
aus
indica
Indica v. group
All varieties
tropical japonica
Japonica v. group

Figure 5.6. Gene linkage and/or pleiotropy surrounding candidate genes. (a) Manhattan plots depicting significant associations for panicle traits in two subpopulations on the distal end of chromosome 10: maximum primary branch length (*tropical japonica* panel), and primary branch number in the inferior half of the panicle (*aus* panel). The x-axis depicts the physical location of SNPs across chromosome 10, and the y-axis depicts the $-\log_{10}(P \text{ value})$. Significant SNPs ($P < 1 \times 10^{-5}$) are depicted as colored dots, labeled to match the group in which they were identified: *tropical japonica* (blue) and *aus* (yellow). The zoom-in plot depicts ~195kb region (grey dotted lines), with gene models in green. The heatmap depicts linkage disequilbrium (LD), measured as r^2, between SNPs calculated using *aus* varieties. The most significant SNP for the *aus* and *tropical japonica* peaks are labeled in yellow and blue. The candidate gene, *WRKY2*, is located ~30kb of the peak SNP for the maximum primary branch length. (b) Manhattan plots depicting significant associations for a grain trait and a panicle trait on chromosome 4: 1000-grain weight (*Indica* subspecies panel), and maximum internode length (*aus* subpopulation panel). Significant SNPs ($P < 1 \times 10^{-5}$) are depicted as colored dots, labeled to match the group in which they were identified: *Indica* (red) and *aus* (yellow). The zoom-in plot depicts ~75kb region (grey dotted lines), with gene models in green. The heatmap depicts linkage disequilbrium (LD), measured as r^2, between SNPs calculated using *aus* varieties. All significant SNPs for *aus* and *Indica* are labeled in yellow and red, respectively. The candidate gene, *OsKS1*, contains two nonsynonymous exonic SNPs (red asterisks).

5.6a


Max Primary Branch Length (xPBL)
tropical japonica
Primary Branch Number Inferior Half (PBNin)
aus
−log10(p)
0
5 Mb
10 Mb
15 Mb
20 Mb
22.934 Mb
23.129 Mb
Chr. 10
22.95 Mb
23 Mb
23.05 Mb
23.1 Mb
WRKY2
r2
1
0

5.6b


1000-Grain Weight (1000GW)
Indica
Max Internode Length (xNL)
aus
−log10(p)
0
5 Mb
10 Mb
15 Mb
20 Mb
25 Mb
30 Mb
35 Mb
30.99 Mb
31.065 Mb
Chr. 4
31 Mb
31.01 Mb
31.02 Mb
31.03 Mb
31.04 Mb
31.05 Mb
31.06 Mb
OsKS1
r2
1
0

As observed in previous studies, several regions of the genome contained significant associations for both panicle architecture and agronomic traits. In some cases, the agronomic traits were vegetative; for example, we identified a peak on chromosome 1 with overlap between panicle length and flag leaf area, and a peak on chromosome 9 for total shoot biomass that overlaps with peaks for many length phenotypes (Figure 5.5b). When mapping for panicle weight, grain number, and 1000 grain weight, which are all metrics of yield performance, we identify associations that overlap with different types of panicle traits: internode number, branch length, and internode length, respectively. Unlike the associations we observed when comparing panicle phenotypes, panicle and agronomic traits never shared exactly the same significant SNPs within a given LD block (Zhao et al., 2011); rather, significant SNPs were often closely linked within the same LD block (<100 kb).

Assessing pleiotropy through candidate gene analyses

A sizeable number of genes involved in rice panicle development have been cloned using molecular analyses and several have been confirmed to pleiotropically affect both panicle and agronomic phenotypes. Having assessed linkage between significant GWAS-SNPs using association networks, we assembled a list of 319 *a priori* candidate genes (Supplementary Table 4), roughly half of which have been characterized molecularly, in order to determine whether any known genes fell within any of our significant peaks. We identified 9 genes located within 30 kb or less of significant associations (roughly 1-3 genes in rice) (Table 5.2). All but one of the candidate genes identified have been linked to hormone signaling cascades. Notably, the three most significant peaks co-localized with candidate genes within the gibberellin signaling cascade and were identified when mapping for agronomic traits: *GID1 (GIBBERELLIN INSENSITIVE*

Table 5.2. Candidate genes identified near significant genome-wide association (GWAS) peaks.

Candidate gene	**Trait**	**Chr.**	**Gene**	**Biological pathway**	**Position[a]**
SD1	Plant height	1	LOC_Os01g66100	Gibberellin enzyme	38,418,739
EP3/LP	Panicle length	2	LOC_Os02g15950	F-box transcription factor; cytokinin homeostasis	9,109,565
OsKS1	Panicle internode length traits	4	LOC_Os04g52230	Gibberellin enzyme	31,029,056
CYP90D3	Panicle branch length, panicle internode length traits	5	LOC_Os05g11130	Brassinosteroid enzyme	6,264,833
GID1	Booting	5	LOC_Os05g33730	Soluble gibberellin receptor	19,891,242
OsGA2 oxidase-5*	Shoot biomass	7	LOC_Os07g01340	Gibberellin enzyme	216,325
OsBZR1	Panicle internode length traits	7	LOC_Os07g39220	Transcription factor; Brassinosteroid homeostasis	23,477,027
FZP	Secondary panicle branching	7	LOC_Os07g47330	AP2 domain transcription factor	28,297,303
WRKY2	Panicle branch length traits	10	LOC_Os10g42850	WRKY transcription factor	23,095,323

a. Closest significant SNP. Position in bp according to MSUv7 assembly. Chr. = chromosome.
*Identified without heading covariate.

DWARF1), a soluble gibberellin receptor (Ueguchi-Tanaka et al., 2005) associated with booting; *SD1* (*SEMIDWARF 1*), a OsGA20-oxidase gibberellin biosynthesis enzyme associated with plant height (Ashikari et al., 2002); and *OsGA2ox-5*, a gibberellin degradation enzyme (Sakamoto et al., 2004) associated with total shoot biomass. As observed in previous studies, significant GWAS-SNPs often appeared next to but not directly within the gene-coding region (Huang et al., 2011).

We detected four transcription factors associated with panicle phenotypes. Two of these genes, *ERECT PANICLE 3* (*EP3,* also known as *LARGER PANICLE*) and *BRASSINAZOLE RESISTANT 1* (*OsBZR1*), were detected when mapping for panicle traits across all varieties in the panel and are transcriptional regulators of the cytokinin and brassinosteroid signaling pathways, respectively (Bai et al., 2007; Piao et al., 2009; Li et al., 2011). The *FRIZZY PANICLE* (*FZP*) gene, a transcriptional regulator of primary branch meristem determinancy (Komatsu, 2003), was detected when mapping across all varieties for the agronomic phenotype secondary branch number. Within *tropical japonica*, we detected a subpopulation specific peak for several primary branch length traits on chromosome 10 (Figure 5.6a; Supplementary Figure 20-22), encompassing the transcription factor *WRKY2* (Ross et al., 2007). The WRKY gene family regulates a range of biological processes in plants, including disease resistance, drought tolerance, and development (Ross et al., 2007; Rushton et al., 2010), and one other member in rice has been linked with panicle exsertion and seed formation (Zhang et al., 2011).

Within the *aus* subpopulation we identified two genes with overlapping associations for different types of traits. The first, *CYP90D3*, was associated with maximum internode length and

maximum length of primary branches in the superior half of the panicle (Supplementary Figure 9, 21). *CYP90D3* is a homolog of *D2* (*DWARF EBISU*), which encodes a cytochrome P450 brassinosteroid biosynthesis enzyme that has been linked with subtle changes in plant architecture and panicle phenotypes (Hong, 2003). A recent study demonstrated that CYP90D2 and CYP90D3 catalyze many of the same reactions *in vitro* (Sakamoto et al., 2012), suggesting that *CYP90D3* may pleiotropically affect multiple panicle phenotypes similar to its homolog. Interestingly, maximum internode length was associated with a second hormone gene in *aus,* rice ent-kaurene synthase (*OsKS1*) (Figure 5.6b). *OsKS1* catalyzes an early step in gibberellin biosynthesis and is tandemly linked with two of its homologs on chromosome 4 (*OsKS2, OsKS3*)(Sakamoto et al., 2004). Using GWAS, we detected five significant genic SNPs within the *OsKS1* gene, two of which are nonsynonymous and located within exons. A significant peak for one thousand grain weight is located ~10 kb downstream, well within the LD block in the region, suggesting that the *OsKS1* gene or *OsKS* gene cluster may pleiotropically affect both panicle and grain characteristics.

5.3 Discussion

Despite the fact that panicles are the grain bearing organs in rice, previous GWAS and QTL studies have characterized panicle traits by collecting a limited number of phenotypes and/or assessing development with plants grown in controlled environments (Zhao et al., 2011; Crowell et al., 2014). Additionally, due to the deep subpopulation structure and extensive LD in *O. sativa*, many studies have performed GWAS in rice using large panels of diverse germplasm in order to increase statistical power and mapping resolution (Huang et al., 2010; Huang et al., 2011; Zhao et al., 2011; Huang et al., 2013). While this approach improves the chance of

detecting significant associations, it considerably confounds panicle trait associations with loci responsible for flowering and ecological adaptation (Atwell et al., 2010; Brachi et al., 2011; Horton et al., 2012). This has left breeders with an incomplete picture of the genetic architecture underlying rice panicle development in adapted lines, and the relationships between panicle traits and yield performance.

Within this study, we demonstrate that with proper controls for subpopulation structure and flowering, it is possible to reduce type I error and identify numerous QTLs associated with reproductive phenotypes. Use of heading date as a covariate within the mixed model, while conservative, eliminated many spurious SNP associations that were presumably confounded with seasonal/ecological adaptation and the geographic origin of the varieties within our panel. Despite statistically rigorous controls and use of an intermediate population size, we were able to identify significant SNPs closely associated with *a priori* candidate genes from pathways known to regulate plant architecture in rice (Table 5.2). The success of this GWAS was undoubtedly due to the combination of our phenotyping resolution and the use of a high quality SNP dataset based on the high-density rice array (McCouch et al, companion paper). These results suggest that for breeders or molecular biologists interested in quantifying the genetic architecture underlying specific traits or aspects of development, medium-size populations can be used to detect both large and small-effect loci in the field (Supplementary Table 3), provided researchers complement dense marker data with precise, well-targeted phenotyping methodologies.

The number of significant associations we detected for panicle traits is consistent with the number of genes identified in microarray studies assessing rice gene expression during

inflorescence development (Furutani et al., 2006; Sato et al., 2011), as well as the number of genes comprising maize inflorescence regulatory modules of gene expression (Eveland et al., 2013). Assessing gene expression during panicle development in rice is difficult, due to the narrow window of time in which meristematic transitions occur (~7 days) and the size of the tissue being sampled (<10mm). Here, we provide evidence that quantitative phenotyping methodologies at a macro level can be used to home in on candidate regions of the genome associated with fine variation in gross morphology. Previous studies assessing flowering time have confirmed that gene expression levels can directly affect trait variation in rice (Takahashi et al., 2009). Our ability to detect distinct associations when mapping for nested traits, especially when separating the panicle axis into inferior and superior halves, suggests that we may be capturing genes with spatiotemporal expression differences during a narrow developmental window. Future studies could assess environmental effects (GxE) on the appearance of these associations in controlled environments and use expression QTL (eQTL) mapping to narrow in on specific genes of interest (Cookson et al., 2009).

Given the strong correlation between panicle traits, we expected that many phenotypes would either share significant SNPs or have associations within similar regions of the genome. While we observed overlapping associations among similar types of panicle traits when mapping across all varieties, SNPs associated with length and count traits often mapped to regions of the genome that were independent of one another. Additionally, when mapping for subpopulation-specific associations, we were surprised that we detected only one region of the genome with overlapping associations for panicle traits in two subpopulations. While subpopulation-specific polymorphisms have been reported for many traits in rice (Huang et al., 2010; Huang et al.,

2011; Zhao et al., 2011; Gamuyao et al., 2013; Huang et al., 2013), the level of variation observed within this study was striking, given that many of the phenotypes we measured share a biological and/or morphological origin. The number of subpopulation-specific panicle associations could be the direct result of the fact that reproductive phenotypes are often closely correlated with ecological adaptation (Atwell et al., 2010; Brachi et al., 2011; Horton et al., 2012) and may explain why different subpopulations have distinct distributions for traits—even in nested phenotypes. Furthermore, the deep, genome-wide differentiation among subpopulations of rice means that mutations in a particular gene often have different phenotypic consequences in different genetic backgrounds (Famoso et al., 2011; Zhao et al., 2011). In cases where a functional mutation or environmental variation perturb a gene that pleiotropically controls multiple aspects of panicle architecture and reproductive behavior, the phenotypic impacts would be particularly noteworthy, in that it could impact the future course of evolutionary possibilities within the lineage through decanalization of otherwise cryptic variation, enhancing reproductive isolation (Flatt, 2005).

Thus, our observations document the quantitative nature of variation for panicle traits among varieties both within and across rice subpopulations, and we hypothesize that combinations of alleles that were not detected using GWAS further enhance the subpopulation-specific morphology described here (Figures 5.1 and 5.2). Our findings also suggest that no one aspect of panicle size has been severely aggrandized or individually optimized during domestication. Rather, panicle morphology appears to be comprised of correlated components that interact and compensate for one another during development, so that selecting for increased size or number of one feature will likely be met with a reduced size or number of a second, correlated trait.

Thus, while the underlying genetics governing panicle architecture in rice are highly subpopulation-specific, the phenotypic outcomes are surprisingly similar. This may also explain why panicles from domesticated rice retain the same essential architecture of those from their wild relatives (Yamaki et al., 2010).

A combination of overlapping and discrete associations between yield components and panicle traits highlights the fact that genetic variation in rice is not only physically partitioned into different subpopulations that maintain their identity through reduced recombination and sterility barriers, but is also spatiotemporally partitioned between developmental transitions (Itoh, 2005; Yoshida and Nagato, 2011). The sheer number of loci contributing to panicle morphology suggest that breeders attempting to improve yield potential in rice may have more success by focusing directly on optimizing grain filling and yield *per se*, rather than by trying to optimize specific features of panicle architecture or panicle development. In keeping with this perspective, we note that the highest yielding *indica* varieties within our panel have characteristically intermediate panicle phenotypes and the smallest average internode length (Figure 5.2), a trait that has been previously linked to yield performance in rice (Huang et al., 2009). However, breeding for specific panicle traits would likely involve managing a highly interactive network or trait complex, and poses a challenge that is as complex as breeding for overall increases in yield itself.

Within the public breeding community, there have been two major initiatives within the past 70 years to boost yield by optimizing independent phases of rice development. The first occurred during the Green Revolution, in which breeders successfully leveraged a large-effect, recessive

allele at the *SD1* locus that optimized vegetative plant architecture without drastically changing reproductive phenotypes (Jennings 1964; Beachell and Jennings 1965; Ashikari et al., 2002). The second initiative, a type of ideotype breeding known as "New Plant Type" (NPT) breeding, attempted to boost yield by simultaneously selecting for increased panicle (sink) size and photosynthetic capacity (source). This was done largely through introgression of QTL associated with large panicle and low-tillering, erect vegetative architecture traits from *tropical japonica* into *indica* varieties (Chen et al., 2001; Khush, 2001; Chen et al., 2007). Given the quantitative nature of panicle development that we detected within and between the *indica* and *tropical japonica* subpopulations, it is not surprising that the NPT initiative ultimately failed to generate high yielding varieties. Integrating multiple loci that coordinately regulate reproductive performance in an effort to optimize yield is a much more difficult problem than selecting for a single gene of large effect that impacts a phenotype at the pre-reproductive stage. Understanding and improving upon traits with highly quantitative inheritance within elite rice germplasm will likely require integration of more sophisticated approaches than the use of traditional marker-assisted selection and introgression. The quantitative tools and strategies used in this study to identify loci related to natural variation for panicle development are primed for integration into genome-assisted crop improvement pipelines. By identifying, understanding, and integrating trait-specific SNP associations into genome wide prediction models, breeders may one day close the gap between panicle development and yield optimization in rice.

Acknowledgements

We would like to thank the International Rice Genebank at the International Rice Research Institute (IRRI) in Los Baños, Philippines, for providing seed stocks; Lovelie Shaine Olivo and

her field crew at IRRI for outstanding technical assistance during the phenotyping field trial; Genevieve DeClerk and Francisco Agosto-Perez at Cornell University (CU) in Ithaca, NY for bioinformatic assistance; Anthony Greenberg at CU for statistical consulting; and Diane Wang at CU for valuable discussion during manuscript preparation. The field trial at IRRI was supported by The Bill and Melinda Gates Foundation, grant awarded to A.I. and G.G. Expansion of the PANorama phenotyping platform was supported by FAPESP grant 2011/03110-6 awarded to A.F. SNP genotype development and data analysis were supported by the NSF Plant Genomes Research Program (Grant #1026555), awarded to S.R.M. S.C. was supported by the NSF Graduate Research Fellowship Program (NSF-GRFP).

Author contributions

S.C. and S.R.M. conceived and designed the experiments. A.I. and G.G. co-supervised the field trial, and S.R.M. supervised the project overall. A.I. and G.G. contributed plant materials and resources surrounding the field trial. S.C. and A.F. conceived new PANorama phenotyping PANorama measurements. A.F. implemented phenotyping measurements into PANorama software. S.C., P.K., J.M., and S.R.M. contributed to data analysis tools. S.C. and P.K. generated the GWAS results. S.C. analyzed population structure, performed phenotyping, assembled the candidate gene list, and developed association network methodology. S.C. and S.R.M. analyzed the data. S.C. and S.R.M. wrote the manuscript. S.C., P.K., A.F., A.I., G.G., J.M., and S.R.M. critically reviewed and approved the manuscript.

Supplemental Materials

The following supplemental data is available in the online version of this article.

Supplementary Figures 1-65. Genome-wide association results.

Supplementary Table 1..Germplasm comprising the diversity panel.

Supplementary Table 2. List of phenotypes used in genome-wide association studies.

Supplementary Table 3. Relative effect sizes (RES) of the most significant SNPs detected for traits within individual subpopulations.

Supplementary Table 4. A prior candidate genes identified within the literature relating to vegetative or reproductive phenotypes.

5.4 Methods

Germplasm selection

A collection of 1,568 accessions representing the five major subpopulations in *O. sativa* was recently genotyped for 700,000 SNPs using a high-density rice array (HDRA) (McCouch et al, companion paper). We wished to maximize the diversity among rice varieties with HDRA genotypes and minimize confounding effects relating to poor adaptation for growth in the tropics. The majority of varieties were selected from three rice subpopulations (63 *aus,* 84 *indica,* 79 *tropical japonica*, 11 admixed *Japonica*, 3 *temperate japonica,* and 2 admixed accessions). Detailed information regarding variety ancestry is located within Supplementary Table 5.1.

Population structure

The principle component analysis was conducted using the svd() function in R[56] (version 3.1.0), calculated using SNPs present in all accessions. The Bayesian clustering program *fastStructure* was used to calculate varying levels of *K* (K=1-10), and the command chooseK.py was used to

identify the model complexity that maximized the marginal likelihood (K=8). Supplementary Figure 1a was generated using the program distruct(Rosenberg, 2004). Genome wide linkage disequilibrium (LD) was estimated using pairwise r^2 between SNPs, which was calculated using the --r2 --ld-window 99999 --ld-window-r2 0 command in PLINK (Purcell et al., 2007) (version 1.07).

Phenotyping details

Three replications of each variety were evaluated during the 2013 dry season (January-May) at the International Rice Research Institute (IRRI) in Los Baños, Philippines in a randomized block design under flooded paddy conditions. Each replication consisted of a two-row plot 4.6 m in length, with 0.2 m between plants and 0.3 m between rows. Panicle traits, heading date, and booting date were collected on all 242 varieties within the panel. All other yield components were collected on 136 randomly sampled varieties from the *indica, aus,* and *tropical japonica* subpopulations (Supplementary Table 1). Detailed descriptions of all phenotypes, acronyms, and measurement methods are presented in (Supplementary Table 2).

Panicle imaging protocol

Following the PANorama imaging protocol(Crowell et al., 2014), 3,443 images were collected and analyzed using a pixel to length conversion of 114.5 pixels/cm. PANorama1.0 contained phenotyping capabilities for 18 major traits, which calculated via image segmentation and subdivision of panicle axes[18]. Additional, nested phenotypes used in this study (i.e. subdivision of the panicle axes into upper and lower halves) were calculated from measurements extracted after the image segmentation and skeletonization process, and thus did not require alternation to

the algorithms implemented in PANorama1.0. Detailed descriptions of these phenotypes are available in Supplementary Table 2. An updated version of PANorama containing all nested phenotypes used within this study, PANorama2.0, is available for download at sourceforge.net/panorama1.

Phenotype statistical analyses

Histograms, boxplots, correlations, and GWAS analyses were constructed using phenotypic grand means for each variety. *P* values for Pearson's correlation coefficients were calculated with a two-sided *t*-test using the cor.test() function in R[56]. To estimate narrow sense (h^2) heritability of a phenotype, the restricted maximum likelihood (REML) estimate of the genetic variance was calculated using the mixed.solve() function in the R package rrBLUP (version 4.3), and the value was divided by the total phenotypic variance. To estimate broad sense heritability (H), the variance among variety grand means was divided by the total phenotypic variance of raw trait values.

Genome-wide association

EMMAX was used to calculate the linear mixed model and significance levels within the GWAS model(Kang et al., 2010). In all GWAS runs, within subpopulations or across all varieties, we used the equation:

$$Y = \alpha X + \beta P + \mu + \varepsilon \quad (1)$$

For GWAS runs incorporating heading date as a covariate:

$$Y = \alpha X + \beta P + \beta HD + \mu + \varepsilon \quad (2)$$

Where Y and X represent the phenotype and SNP genotype vectors, respectively; P is a matrix containing the residuals of the first 3 principle components (PCs); and HD represents a vector of the heading date phenotype. For genotypic and environmental random effects, respectively, $u \sim N(0, \sigma^2_g K)$ and $\varepsilon \sim N(0, \sigma^2_e I)$, where K is an identity by state (IBS) kinship matrix accounting for pairwise relatedness between varieties. SNP marker filtering (minor allele frequency = 0.1, genetic missingness = 0.3) and IBS matrix calculations were performed using PLINK (Purcell et al., 2007). Because yield components were collected on a subset of the varieties within our panel (Supplementary Table 1), we performed GWAS for these traits within the *Indica* and *Japonica* subspecies rather than in the individual subpopulations in order to maximize our power to detect loci. We noted that certain traits were more susceptible to confounding than others, especially when performing GWAS across subpopulations using the entire panel of varieties. In order to correct for these issues, we systematically diagnosed the quantile-quantile plots for every trait-subpopulation-covariate combination and used logarithmic transformations on non-normal phenotypes (Supplementary Table 2). The significance threshold was set at $P < 1 \times 10^{-5}$ for every trait, in accordance with (McCouch et al, companion paper).

Relative effect size

For each trait, the relative effect size (Supplementary Table 3) was calculated for the most significant SNP detected within each subpopulation using the following equation:

$$\frac{n_{Major}*(x_{Major} - x)^2 + n_{Minor}*(x_{Minor} - x)^2}{\sum_{i=1}^{n}(x_i - x)^2} \quad (3)$$

Where n is the total number of individuals; n_{Major} and n_{Minor} are the number of individuals with the major and minor alleles, respectively; x_{Major} and x_{Minor} are the trait means for individuals with

the major and minor alleles, respectively; x_i is the trait value of individual i; and x is the trait mean of the n individuals with values x_i.

Association networks

Significant SNPs were binned together into peaks using a sliding window based off a linkage disequilbrium using the PLINK (Purcell et al., 2007) command --clump-p1 0.00001 --clump-p2 0.0001 --clump-r2 0.3 --clump-kb 150 --clump-allow-overlap. Thus, for every SNP with $P < 1 \times 10^{-5}$, pairwise r^2 values were calculated between surrounding SNPs that 1) fell within 150kb, and 2) had a $P < 1 \times 10^{-3}$; any two SNPs meeting this criteria that also shared an $r^2 \geq 0.3$ were clumped into bins. All significant SNPs within the study were used in the construction of bins, regardless of the traits with which they shared associations. Additionally, any bins sharing overlapping borders after using the PLINK clump command were collapsed into a single bin. Singleton, significant SNPs ($< 1 \times 10^{-5}$) were discarded if no other SNP within the LD window was $< 2.5 \times 10^{-4}$. In order to construct association networks, traits and their corresponding bins were treated as nodes within the program Cytoscape(Shannon et al., 2003) (version 3.1), and edges were labeled by the subpopulation in which the trait association was identified.

Candidate gene analyses

A list of 319 candidate genes was assembled using a literature review and BLAST searches for candidate gene homologs (Supplementary Table 4). LD plots and r^2 values for candidate gene zoom-ins were constructed using Haploview (Barrett et al., 2005) (version 4.2).

REFERENCES

3,000 rice genomes project (2014) The 3,000 rice genomes project. Gigascience **3**: 7

Alston JM, Pardey PG, and Beddow JM (2010) Global Patterns of Crop Yields and Other Partial Productivity Measures and Prices. In *The Shifting Patterns of Agricultural Production and Productivity Worldwide*, ed. J. M. Alston, B. Babcock and P. G. Pardey, Chapter 3. Ames: Iowa State University

AL-Tam FM, Adam H, Anjos AD, Lorieux M, Larmande P, Ghesquière A, Jouannic S, Shahbazkia HR (2013) P-TRAP: a Panicle Trait Phenotyping Tool. BMC Plant Biol **13**: 122

Amrhein N, Deus B, Gehrke P, Steinrücken HC (1980) The Site of the Inhibition of the Shikimate Pathway by Glyphosate: II. INTERFERENCE OF GLYPHOSATE WITH CHORISMATE FORMATION IN VIVO AND IN VITRO. Plant Physiology **66**: 830–834

Ando T, Yamamoto T, Shimizu T, Ma XF, Shomura A, Takeuchi Y, Lin SY, Yano M (2008) Genetic dissection and pyramiding of quantitative traits for panicle architecture by using chromosomal segment substitution lines in rice. TAG Theoretical and Applied Genetics **116**: 881–890

Araus JL, Cairns JE (2014) Field high-throughput phenotyping: the new crop breeding frontier. Trends in Plant Science **19**: 52–61

Asano K, Yamasaki M, Takuno S, Miura K, Katagiri S, Ito T, Doi K, Wu J, Ebana K, Matsumoto T (2011) Artificial selection for a green revolution gene during japonica rice domestication. Proceedings of the National Academy of Sciences **108**: 11034

Ashikari M (2005) Cytokinin Oxidase Regulates Rice Grain Production. Science **309**: 741–745

Ashikari M, Sasaki A, Ueguchi-Tanaka M, Itoh H, Nishimura A, Datta S, Ishiyama K, Saito T, Kobayashi M, Khush GS (2002) Loss-of-function of a rice gibberellin biosynthetic gene, GA20 oxidase (GA20ox-2), led to the rice"green revolution." Breed Sci **52**: 143–150

Atwell S, Huang YS, Vilhjálmsson BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, Hu TT, et al (2010) Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature **465**: 627–631

Bai M-Y, Zhang L-Y, Gampala SS, Zhu S-W, Song W-Y, Chong K, Wang Z-Y (2007) Functions of OsBZR1 and 14-3-3 proteins in brassinosteroid signaling in rice. Proceedings of the National Academy of Sciences **104**: 13839–13844

Barrett CB, Reardon T, Webb P (2001) Nonfarm income diversification and household livelihood strategies in rural Africa: concepts, dynamics, and policy implications. Food

Policy **26**: 315–331

Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics **21**: 263–265

Bates SL, Zhao J-Z, Roush RT, Shelton AM (2005) Insect resistance management in GM crops: past, present and future. Nat Biotechnol **23**: 57–62

Beachell HM, Jennings PR (1965) Need for modification of plant type. In: The Mineral Nutrition of the Rice Plant, International Rice Research Institute. Baltimore: John Hopkins Press.

Bhullar NK, Street K, Mackay M, Yahiaoui N, Keller B (2009) Unlocking wheat genetic resources for the molecular identification of previously undescribed functional alleles at the Pm3 resistance locus. Proceedings of the National Academy of Sciences **106**: 9519–9524

Bigirimana J, Njoroge K, Gahakwa D, Phiri NA (2012) Incidence and severity of coffee leaf rust and other coffee pests and diseases in Rwanda. African Journal of Agricultural Research **7**: 3847–3852

Bloss R (2014) Robot innovation brings to agriculture efficiency, safety, labor savings and accuracy by plowing, milking, harvesting, crop tending/picking and monitoring. Industrial Robot **41**: 493–499

Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, Bonas U (2009) Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. Science **326**: 1509–1512

Bolot S, Abrouk M, Masood-Quraishi U, Stein N, Messing J, Feuillet C, Salse J (2009) The "inner circle" of the cereal genomes. Current Opinion in Plant Biology **12**: 119–125

Boonekamp P, Haverkort A, Hutten R, Jacobsen E, Lotz B, Kessel G, Visser R, and van der Vossen E (2008) Sustainable resistance against *Phytophthora* in potato through cisgenic marker-free modification. Wageningen, The Netherlands: Project Group DuRPh, Plant Research International

Brachi B, Morris GP, Borevitz JO (2011) Genome-wide association studies in plants: the missing heritability is in the field. Genome Biol **12**: 232

Brar DS and Khush GS (2003) Utilization of wild species of genus *Oryza* in rice improvement. In *Monograph on genus Oryza,* ed. J. S. Nanda and S. D. Sharma, 283–309. Enfield, New Hampshire and Plymouth, UK: Science Publishers

Brookes G, Barfoot P (2010) Global impact of biotech crops: Environmental effects, 1996-2008.

Byerlee D, de Janvry A, Sadoulet E (2009) Agriculture for Development: Toward a New

Paradigm. Annu Rev Resour Econ **1**: 15–31

Cambia BiOS Licensing. Accessed May 2012. www.bios.net/daisy/bios/mta/license-intro.html

Chen LY, Xiao YH, Tang WB, Lei DY (2007) Practices and prospects of super hybrid rice breeding. Rice Science **14**: 71–77

Chen W-F, Xu Z-J, Zhang W-Z, Zhang L-B, Yang S-R (2001) Creation of New Plant Type and Breeding Rice for Super High Yield. Acta Agronomica Sinica

Cheng S-H, Zhuang J-Y, Fan Y-Y, Du J-H, Cao L-Y (2007) Progress in research and development on hybrid rice: a super-domesticate in China. Annals of Botany **100**: 959–966

Choudhary B and Gaur K (2009) The development and regulation of Bt brinjal in India (eggplant/aubergine). ISAAA Brief No. 38–2009. http://www.isaaa.org/resources/publications/briefs/38/default.asp. Accessed June 6, 2012

Clark RT, Famoso AN, Zhao K, SHAFF JE, CRAFT EJ, Bustamante CD, Mccouch SR, ANESHANSLEY DJ, Kochian LV (2012) High-throughput two-dimensional root system phenotyping platform facilitates genetic analysis of root growth and development. Plant, Cell & Environment **36**: 454–466

Clark RT, MacCurdy RB, Jung JK, Shaff JE, McCouch SR, Aneshansley DJ, Kochian LV (2011) Three-Dimensional Root Phenotyping with a Novel Imaging and Software Platform. Plant Physiology **156**: 455–465

Cobb JN, Declerck G, Greenberg A, Clark R, McCouch S (2013) Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype–phenotype relationships and its relevance to crop improvement. Theor Appl Genet **126**: 867–887

Cookson W, Liang L, Abecasis G, Moffatt M, Lathrop M (2009) Mapping complex disease traits with global gene expression. Nature Reviews Genetics **10**: 184–194

Crowell S, Falcão AX, Shah A, Wilson Z, Greenberg AJ, Mccouch SR (2014) High-Resolution Inflorescence Phenotyping Using a Novel Image Analysis Pipeline, PANorama. Plant Physiology. doi: 10.1104/pp.114.238626

Day PR (1997) Biodiversity and the equitable use of the world's genetic resources. In *Global genetic resources: Access, ownership and intellectual property rights*, ed. K. E. Hoagland and A. Y. Rossman. Beltsville symposia in Agricultural Research. Washington, DC: The Association of Systematics Collections

Day-Rubenstein KA, Heisey PW, Shoemaker RA, Sullivan J, Frisvold GB (2005) Crop genetic resources: an economic appraisal.

Delp CJ (1980) Coping with resistance to plant disease. Plant Dis **64**: 652–657

Dita MA, Waalwijk C, Buddenhagen IW, Souza MT Jr, Kema GHJ (2010) A molecular diagnostic for tropical race 4 of the banana fusarium wilt pathogen. Plant Pathology **59**: 348–357

Donnelly JS (2001) *The Great Irish Potato Famine*. Gloucestershire, UK: Sutton Publishing

Douches DS, Maas D, Jastrzebski K, Chase RW (1996) Assessment of potato breeding progress in the USA over the last century. Crop Science **36**: 1544–1552

Doust A (2007) Architectural Evolution and its Implications for Domestication in Grasses. Annals of Botany **100**: 941–950

Duvick DN (1991) Industry and its role in plant diversity. *Forum for Applied Research and Public Policy* **6**:3

Eathington SR, Crosbie TM, Edwards MD, Reiter RS, Bull JK (2007) Molecular Markers in a Commercial Breeding Program. Crop Science **47**: S–154

Eveland AL, Goldshmidt A, Pautler M, Morohashi K, Liseron-Monfils C, Lewis MW, Kumari S, Hiraga S, Yang F, Unger-Wallace E, et al (2013) Regulatory modules controlling maize inflorescence architecture. Genome Research. doi: 10.1101/gr.166397.113

Falcao AX, da Fontoura Costa L, Da Cunha BS (2002) Multiscale skeletons by image foresting transform and its application to neuromorphometry. Pattern recognition **35**: 1571–1582

Falcão AX, Stolfi J, de Alencar Lotufo R (2004) The image foresting transform: theory, algorithms, and applications. IEEE Trans Pattern Anal Mach Intell **26**: 19–29

Famoso AN, Clark RT, Shaff JE, Craft E, McCouch SR, Kochian LV (2010) Development of a Novel Aluminum Tolerance Phenotyping Platform Used for Comparisons of Cereal Aluminum Tolerance and Investigations into Rice Aluminum Tolerance Mechanisms. Plant Physiology **153**: 1678–1691

Famoso AN, Zhao K, Clark RT, Tung CW, Wright MH, Bustamante C, Kochian LV, McCouch SR (2011) Genetic architecture of aluminum tolerance in rice (Oryza sativa) determined through genome-wide association analysis and QTL mapping. PLoS Genet **7**: e1002221

FAO (2010) Food Outlook. Rome: Trade and Markets Division, FAO

FAO (2012) Conservation Agriculture. Rome: FAO. http://www.fao.org/ag/ca/1a.html. Accessed June 5, 2012

Feltus FA, Wan J, Schulze SR, Estill JC, Jiang N, Paterson AH (2004) An SNP resource for rice genetics and breeding based on subspecies indica and japonica genome alignments. Genome Research **14**: 1812–1819

Ferroni M, Castle P (2011) Public-Private Partnerships and Sustainable Agricultural Development. Sustainability **3**: 1064–1073

Flatt T (2005) The Evolutionary Genetics of Canalization. The Quarterly Review of Biology **80**: 287–316

Fukao T, Harris T, Bailey-Serres J (2008) Evolutionary analysis of the Sub1 gene cluster that confers submergence tolerance to domesticated rice. Annals of Botany **103**: 143–150

Furutani I, Sukegawa S, Kyozuka J (2006) Genome-wide analysis of spatial and temporal gene expression in rice panicle development. The Plant Journal **46**: 503–511

Gamuyao R, Chin JH, Pariasca-Tanaka J, Pesaresi P, Catausan S, Dalid C, Slamet-Loedin I, Tecson-Mendoza EM, Wissuwa M, Heuer S (2013) The protein kinase Pstol1 from traditional rice confers tolerance of phosphorus deficiency. Nature **488**: 535–539

Generation Challenge Program (2012) The GCP Molecular Marker Toolkit. http://www.generationcp.org/sp5/?da=10179508. Accessed August 13, 2012

Gao S, Yu B, Yuan L, Zhai H, He S-Z, Liu Q-C (2011) Scientia Horticulturae. Scientia Horticulturae **128**: 408–414

Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S (2005) Genetic Structure and Diversity in Oryza sativa L. Genetics **169**: 1631–1638

Gepts P (2002) A comparison between crop domestication, classical plant breeding, and genetic engineering. Crop Science **42**: 1780

Goff SA (2002) A Draft Sequence of the Rice Genome (Oryza sativa L. ssp. japonica). Science **296**: 92–100

Gouache JC (2004) Balancing access and protection: Lessons from the past to build the future. ISF International Seminar on the Protection of Intellectual Property and Access to Plant Genetic Resources, Berlin, Germany 27-28 May, 2004

Greenberg AJ, Hackett SR, Harshman LG, Clark AG (2011) Environmental and genetic perturbations reveal different networks of metabolic regulation. Mol Syst Biol **7**: 563

Hammond E (2010) The sorghum gene grab. A briefing paper by the African Centre for Biosafety. June 2010. http://www.acbio.org.za/index.php/publications/biopiracy/314-the-sorghum-gene-grab. Accessed August 13, 2012

Hattori Y, Nagai K, Furukawa S, Song X-J, Kawano R, Sakakibara H, Wu J, Matsumoto

T, Yoshimura A, Kitano H, et al (2009) The ethylene response factors SNORKEL1 and SNORKEL2 allow rice to adapt to deep water. Nature **460**: 1026–1030

Hedden P (2003) The genes of the green revolution. Trends in Genetics **19**: 5–9

Heffner EL, Sorrells ME, Jannink J-L (2009) Genomic Selection for Crop Improvement. Crop Science **49**: 1

Heller MA, Eisenberg RS (1998) Can patents deter innovation? The anticommons in biomedical research. Science **280**: 698–701

Herring RJ (2007) Stealth seeds: Bioproperty, biosafety, biopolitics. Journal of Development Studies **43**: 130–157

Herring RJ, Rao NC (2012) On the "Failure of Bt Cotton." Economic & Political Weekly **47**: 45

Hittalmani S, Huang N, Courtois B, Venuprasad R, Shashidhar HE, Zhuang JY, Zheng KL, Liu GF, Wang GC, Sidhu JS, et al (2003) Identification of QTL for growth- and grain yield-related traits in rice across nine locations of Asia. Theor Appl Genet **107**: 679–690

Hobbs PR (2007) Conservation agriculture: what is it and why is it important for future sustainable food production? JOURNAL OF AGRICULTURAL SCIENCE-CAMBRIDGE- **145**: 127

Hong Z (2003) A Rice Brassinosteroid-Deficient Mutant, ebisu dwarf (d2), Is Caused by a Loss of Function of a New Member of Cytochrome P450. The Plant Cell Online **15**: 2900–2910

Hope J (2008) *Biobazaar: The open source revolution and biotechnology*. Cambridge, MA: Harvard University Press

Hope J (2009) Patent pools, clearinghouses, open source models and liability regimes. In *Gene patents and collaborative licensing models*, ed. G. Van Overwalle. Cambridge, MA: Cambridge University Press.

Horton MW, Hancock AM, Huang YS, Toomajian C, Atwell S, Auton A, Muliyati NW, Platt A, Sperone FG, Vilhjálmsson BJ, et al (2012) Genome-wide patterns of genetic variation in worldwide Arabidopsis thaliana accessions from the RegMap panel. Nature Publishing Group **44**: 212–216

Huang X, Kurata N, Wei X, Wang Z-X, Wang A, Zhao Q, Zhao Y, Liu K, Lu H, Li W, et al (2013) A map of rice genome variation reveals the origin of cultivated rice. Nature **490**: 497–501

Huang X, Qian Q, Liu Z, Sun H, He S, Luo D, Xia G, Chu C, Li J, Fu X (2009) Natural variation at the DEP1 locus enhances grain yield in rice. Nat Genet **41**: 494–497

Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, et al (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. Nature Publishing Group **42**: 961–967

Huang X, Zhao Y, Wei X, Li C, Wang A, Zhao Q, Li W, Guo Y, Deng L, Zhu C, et al (2011) Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. Nat Genet 1–10

Ikeda K, Sunohara H, Nagato Y (2004) Developmental course of inflorescence and spikelet in rice. Breed Sci **54**: 147–156

Ikeda M, Hirose Y, Takashi T, Shibata Y, Yamamura T, Komura T, Doi K, Ashikari M, Matsuoka M, Kitano H (2010) Analysis of rice panicle traits and detection of QTLs using an image analyzing method. Breed Sci **60**: 55–64

International Plant Genetic Resources Institute, Association WARD (2007) Descriptors for wild and cultivated Rice (Oryza spp.).

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature **436**: 793–800

ISAAA (2011) Global status of commercialized biotech/GM crops: Brief No.43-2011. http://www.isaaa.org/resources/publications/briefs/43/. Accessed June 5, 2012

ISAAA (2013) Global status of commercialized biotech/GM crops: Brief No.46-2013. http://www.isaaa.org/resources/publications/briefs/46/executivesummary/. Accessed November 4, 2014

Ishii T, Numaguchi K, Miura K, Yoshida K, Thanh PT, Htun TM, Yamasaki M, Komeda N, Matsumoto T, Terauchi R, et al (2013) OsLG1 regulates a closed panicle trait in domesticated rice. Nature Publishing Group **45**: 462–5– 465e1–2

Itoh JI (2005) Rice Plant Development: from Zygote to Spikelet. Plant and Cell Physiology **46**: 23–47

Iyer-Pascuzzi AS, Symonova O, Mileyko Y, Hao Y, Belcher H, Harer J, Weitz JS, Benfey PN (2010) Imaging and Analysis Platform for Automatic Phenotyping and Trait Ranking of Plant Root Systems. Plant Physiology **152**: 1148–1157

Jacobsen E, Schouten HJ (2008) Cisgenesis, a New Tool for Traditional Plant Breeding, Should be Exempted from the Regulation on Genetically Modified Organisms in a Step by Step Approach. Potato Res **51**: 75–88

Jennings PR (1964) Plant type as a rice breeding objective. Crop Sci. **4**:13-15

Kang HM, Sul JH, Service SK, Zaitlen NA, Kong S-Y, Freimer NB, Sabatti C, Eskin E

(2010) Variance component model to account for sample structure in genome-wide association studies. Nature Publishing Group **42**: 348–354

Kathage J, Qaim M (2012) Economic impacts and impact dynamics of Bt (Bacillus thuringiensis) cotton in India. Proceedings of the National Academy of Sciences **109**: 11652–11656

Kellogg EA (2007) Floral displays: genetic control of grass inflorescences. Current Opinion in Plant Biology **10**: 26–31

Khush GS (2001) Green revolution: the way forward. Nature Reviews Genetics **2**: 815–822

Klee H, Horsch R, Rogers S (1987) Agrobacterium-mediated plant transformation and its further applications to plant biology. Annual review of plant physiology **38**: 467–486

Kobayashi S, Fukuta Y, Sato T, Osaki M, Khush GS (2003) Molecular marker dissection of rice (Oryza sativa L.) plant architecture under temperate and tropical climates. Theor Appl Genet **107**: 1350–1356

Kolady D, Spielman DJ, Cavalieri AJ (2010) Intellectual property rights, private investment in research, and productivity growth in Indian agriculture: A review of evidence and options.

Komatsu M (2003) FRIZZY PANICLE is required to prevent the formation of axillary meristems and to establish floral meristem identity in rice spikelets. Development **130**: 3841–3850

Krishna VV, Qaim M (2008) Potential impacts of Bt eggplant on economic surplus and farmers' health in India. Agricultural Economics **38**: 167–180

Kronstad WE (1996) Genetic diversity and the free exchange of germplasm in breaking yield barriers. In *Increasing Yield Potential in Wheat: Breaking the Barriers*, ed. M.P. Reynolds, S. Rajaram, and A. McNab. Mexico, 19–27

Latif MA, Islam MR, Ali MY, Saleque MA (2005) Validation of the system of rice intensification (SRI) in Bangladesh. Field Crops Research **93**: 281–292

Lerner J and Schankerman M (2010) *The conmingled code: Open source and economic development*. Cambridge, MA: MIT Press

Li M, Tang D, Wang K, Wu X, Lu L, Yu H, Gu M, Yan C, Cheng Z (2011) Mutations in the F-box gene LARGER PANICLE improve the panicle architecture and enhance the grain yield in rice. Plant Biotechnology Journal no–no

Li T, Liu B, Spalding MH, Weeks DP, Yang B (2012) High-efficiency TALEN-based gene editing produces disease-resistant rice. Nat Biotechnol **30**: 390–392

Li ZK, Yu SB, Lafitte HR, Huang N, Courtois B, Hittalmani S, Vijayakumar CHM, Liu GF, Wang GC, Shashidhar HE, et al (2003) QTL x environment interactions in rice. I. heading date and plant height. TAG Theoretical and Applied Genetics **108**: 141–153

Liu GL, Mei HW, Yu XQ, Zou GH, Liu HY, Hu SP, Li MS, Wu JH, Chen L, Luo LJ (2008) QTL analysis of panicle neck diameter, a trait highly correlated with panicle size, under well-watered and drought conditions in rice (Oryza sativa L.). Plant Science **174**: 71–77

Luo A, Qian Q, Yin H, Liu X, Yin C, Lan Y, Tang J, Tang Z, Cao S, Wang X, et al (2006) EUI1, encoding a putative cytochrome P450 monooxygenase, regulates internode elongation by modulating gibberellin responses in rice. Plant and Cell Physiology **47**: 181–191

Mali P, Esvelt KM, Church GM (2013) Cas9 as a versatile tool for engineering biology. Nat Meth **10**: 957–963

Merges R (2001) Institutions for intellectual property transactions: The case of patent pools. In *Expanding the Boundaries of intellectual property*, ed. R. Dreyfuss, D. Leenheer Zimmerman, and H. First, 123–66. Oxford: Oxford University Press

McDonald BA, Linde C (2002a) The population genetics of plant pathogens and breeding strategies for durable resistance. Euphytica **124**: 163–180

McDonald BA, Linde C (2002b) Pathogen population genetics, evolutionary potential, and durable resistance. Annu Rev Phytopathol **40**: 349–379

McNally KL, Childs KL, Bohnert R, Davidson RM, Zhao K, Ulat VJ, Zeller G, Clark RM, Hoen DR, Bureau TE, et al (2009) Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. Proceedings of the National Academy of Sciences of the United States of America **106**: 12273–12278

McSteen P (2006) Branching out: the ramosa pathway and the evolution of grass inflorescence morphology. Plant Cell **18**: 518–522

Mulla DJ (2013) Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. Biosystems Engineering **114**: 358–371

Mundt CC (2002) Use of multiline cultivars and cultivar mixtures for disease management. Annu Rev Phytopathol **40**: 381–410

Naqvi S, Zhu C, Farre G, Ramessar K, Bassie L, Breitenbach J, Perez Conesa D, Ros G, Sandmann G, Capell T (2009) Transgenic multivitamin corn through biofortification of endosperm with three vitamins representing three distinct metabolic pathways. Proceedings of the National Academy of Sciences **106**: 7762

Nishiguchi M, Mori M, Okada Y, Murata T, Kimura T, Sakai JI, Hanada K, Miyazaki C, Saito A (1998) Virus resistant transgenic sweet potato with the CP gene: Current challenge

and perspective of its use. Phytoprotection **79**: 112–116

Ohsumi A, Takai T, Ida M, Yamamoto T, Arai-Sanoh Y, Yano M, Ando T, Kondo M (2011) Evaluation of yield performance in rice near-isogenic lines with increased spikelet number. Field Crops Research **120**: 68–75

Palloix A, Ayme V, Moury B (2009) Durability of plant major resistance genes to pathogens depends on the genetic background, experimental evidence and consequences for breeding strategies. New Phytologist **183**: 190–199

Pardey PG, Koo B, and Nottenburg C (2004) Creating, protecting and using crop biotechnologies worldwide in an era of intellectual property. In: WIPO/UPOV Symposium on Intellectual Property Rights in Plant Biotechnology. Geneva, October 10, 2003. WIPO: Geneva

Pardey PG, Beintema NM, Dehmer S, Wood S (2006) Agricultural Research: A growing global divide? 17:

Pardey PG, Pingali P (2010) Reassessing International Agricultural Research for Food and Agriculture. 28–31

Piao R, Jiang W, Ham T-H, Choi M-S, Qiao Y, Chu S-H, Park J-H, Woo M-O, Jin Z, An G, et al (2009) Map-based cloning of the ERECT PANICLE 3 gene in rice. TAG Theoretical and Applied Genetics **119**: 1497–1506

Powles SB (2008) Evolved glyphosate-resistant weeds around the world: lessons to be learnt. Pest Manag Sci **64**: 360–365

Prakash C, Varadarajan U (1992) Genetic transformation of sweet potato by particle bombardment. Plant Cell Rep **11**: 53–57

ProVitaMinRice Consortium (2012) Grand Challenges in Global Health. http://www.goldenrice.org/Content5-GCGH/GCGH1.html. Accessed June 5, 2012

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al (2007) PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. The American Journal of Human Genetics **81**: 559–575

Que Q, Chilton M-DM, de Fontes CM, He C, Nuccio M, Zhu T, Wu Y, Chen JS, Shi L (2010) Trait stacking in transgenic crops: Challenges and opportunities. gmcrops **1**: 220–229

Raj A, Stephens M, Pritchard JK (2014) fastSTRUCTURE: Variational Inference of Population Structure in Large SNP Data Sets. Genetics **197**: 573–589

Regal PJ (1996) Metaphysics in genetic engineering: Cryptic philosophy and ideology in the

"science of risk assessment". In C*oping with deliberate release: The limits of risk assessment*, ed. Ad Van Dommelen, 15–32. International Centre for Human and Public Affairs, Tilburg/Buenos Aires

Rosenberg NA (2004) distruct: a program for the graphical display of population structure. Molecular Ecology Notes **4**: 137–138

Ross CA, Liu Y, Shen QJ (2007) The WRKY gene family in rice (Oryza sativa). J Integr Plant Biol **49**: 827–842

Rushton PJ, Somssich IE, Ringler P, Shen QJ (2010) WRKY transcription factors. Trends in Plant Science **15**: 247–258

Sakai H, Kanamori H, Arai-Kichise Y, Shibata-Hatta M, Ebana K, Oono Y, Kurita K, Fujisawa H, Katagiri S, Mukai Y, et al (2014) Construction of Pseudomolecule Sequences of the aus Rice Cultivar Kasalath for Comparative Genomics of Asian Cultivated Rice. DNA Research **21**: 397–405

Sakai H, Lee SS, Tanaka T, Numa H, Kim J, Kawahara Y, Wakimoto H, Yang CC, Iwamoto M, Abe T, et al (2013) Rice Annotation Project Database (RAP-DB): An Integrative and Interactive Database for Rice Genomics. Plant and Cell Physiology **54**: e6–e6

Sakamoto T, Miura K, Itoh H, Tatsumi T, Ueguchi-Tanaka M, Ishiyama K, Kobayashi M, Agrawal GK, Takeda S, Abe K, et al (2004) An overview of gibberellin metabolism enzyme genes and their related mutants in rice. Plant Physiology **134**: 1642–1653

Sakamoto T, Ohnishi T, Fujioka S, Watanabe B, Mizutani M (2012) Rice CYP90D2 and CYP90D3 catalyze C-23 hydroxylation of brassinosteroids in vitro. Plant Physiol Biochem **58**: 220–226

Sato Y, Antonio B, Namiki N, Motoyama R, Sugimoto K, Takehisa H, Minami H, Kamatsuki K, Kusaba M, HIROCHIKA H, et al (2011) Field transcriptome revealed critical developmental and physiological transitions involved in the expression of growth potential in japonica rice. BMC Plant Biol **11**: 10

Sato Y, Namiki N, Takehisa H, Kamatsuki K, Minami H, Ikawa H, Ohyanagi H, Sugimoto K, Itoh JI, Antonio BA, et al (2012) RiceFREND: a platform for retrieving coexpressed gene networks in rice. Nucleic Acids Research **41**: D1214–D1221

Schatz MC, Maron LG, Stein JC, Wences AH, Gurtowski J, Biggers E, Lee H, Kramer M, Antoniou E, Ghiban E (2014) Whole genome de novo assemblies of three divergent strains of rice, Oryza sativa, document novel gene space of aus and indica. Genome Biol **15**: 506

Schubert D, Williams D (2006) "Cisgenic" as a product designation. Nat Biotechnol **24**: 1327–9– author reply 1331–3

Second G (1982) Origin of the genic diversity of cultivated rice (Oryza spp.): study of the polymorphism scored at 40 isozyme loci. Jpn J genet **57**: 25–57

Septiningsih EM, Pamplona AM, Sanchez DL, Neeraja CN, Vergara GV, Heuer S, Ismail AM, Mackill DJ (2008) Development of submergence-tolerant rice cultivars: the Sub1 locus and beyond. Annals of Botany **103**: 151–160

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Research **13**: 2498–2504

Shapiro C (2001) Navigating the patent thicket: Cross licenses, patent pools, and standard-setting. In *Innovation Policy and the Economy*. ed. I. Jaffe and B. Adam et al., 119–50. Cambridge: MIT Press

Shelton AM (2012) Genetically engineered vegetables expressing proteins from Bacillus thuringiensis for insect resistance: successes, disappointments, challenges and ways to move forward. GM Crops Food **3**: 175–183

Shuette J (1998) Environmental fate of glyphosate. Environmental Monitoring & Pest Management Department of Pesticide Regulation Sacramento, CA. www.cdpr.ca.gov/docs/emon/pubs/fatememo/glyphos.pdf

Shull GH (1909) A pure-line method in corn breeding. Journal of Heredity 51–58

Singh RP, Hodson DP, Huerta-Espino J, Jin Y, Bhavani S, Njau P, Herrera-Foessel S, Singh PK, Singh S, Govindan V (2011) The Emergence of Ug99 Races of the Stem Rust Fungus is a Threat to World Wheat Production. Annu Rev Phytopathol **49**: 465–481

Smolders W (2005) Plant Genetic Resources for Food and Agriculture: Facilitated Access or Utility Patents on Plant Varieties. IP Strategy Today 1–17

Spindel J, Wright M, Chen C, Cobb J, Gage J, Harrington S, Lorieux M, Ahmadi N, McCouch S (2013) Bridging the genotyping gap: using genotyping by sequencing (GBS) to add high-density SNP markers and new value to traditional bi-parental mapping and breeding populations. Theor Appl Genet **126**: 2699–2716

Sprankle P, Meggitt W, Penner D (1975) Adsorption, mobility, and microbial degradation of glyphosate in the soil. Weed Science 229–234

Steinrücken HC, Amrhein N (1980) The herbicide glyphosate is a potent inhibitor of 5-enolpyruvyl-shikimic acid-3-phosphate synthase. Biochemical and Biophysical Research Communications **94**: 1207–1212

Tabashnik BE, Van Rensburg J, Carrière Y (2009) Field-evolved insect resistance to Bt

crops: definition, theory, and data. Journal of economic entomology **102**: 2011–2025

Takahashi Y, Teshima KM, Yokoi S, Innan H, Shimamoto K (2009) Variations in Hd1 proteins, Hd3a promoters, and Ehd1 expression levels contribute to diversity of flowering time in cultivated rice. Proceedings of the National Academy of Sciences of the United States of America **106**: 4555–4560

Tanabata T, Shibaya T, Hori K, Ebana K, Yano M (2012) SmartGrain: High-Throughput Phenotyping Software for Measuring Seed Shape through Image Analysis. Plant Physiology **160**: 1871–1880

Tatum LA (1971) The southern corn leaf blight epidemic. Science **171**: 1113–1116

Tennant P, Gonsalves C, Ling K, Fitch M, Manshardt R, Slightom J, Gonsalves D (1994) Differential protection against papaya ringspot virus isolates in coat protein gene transgenic papaya and classically cross-protected papaya. Phytopathology **84**: 1359–1365

Thakur AK, Rath S, Roychowdhury S, Uphoff N (2010) Comparative Performance of Rice with System of Rice Intensification (SRI) and Conventional Management using Different Plant Spacings. Journal of Agronomy and Crop Science **196**: 146–159

The Treaty (2004) International Treaty on Plant Genetic Resources for Food and Agriculture. http://www.planttreaty.org/

Troyer AF (1999) Background of US hybrid corn. Crop Science **39**: 601–626

Troyer AF (2004) Background of US Hybrid Corn II. Crop Science **44**: 370–380

Ueguchi-Tanaka M, Ashikari M, Nakajima M, Itoh H, Katoh E, Kobayashi M, Chow T-Y, Hsing Y-IC, Kitano H, Yamaguchi I, et al (2005) GIBBERELLIN INSENSITIVE DWARF1 encodes a soluble receptor for gibberellin. Nature **437**: 693–698

Ullstrup AJ (1972) The impacts of the southern corn leaf blight epidemics of 1970-1971. Annu Rev Phytopathol **10**: 37–50

UPOV (2002) Union for the protection of new varieties of plants (UPOV). http://www.upov.org/resource/en/introduction_dus.html

van Wordragen MF, Dons HJM (1992) Agrobacterium tumefaciens-mediated transformation of recalcitrant crops. Plant Mol Biol Rep **10**: 12–36

Vanblaere T, Szankowski I, Schaart J, Schouten H, Flachowsky H, Broggini GAL, Gessler C (2011) Journal of Biotechnology. Journal of Biotechnology **154**: 304–311

Wang Y, Cheng X, Shan Q, Zhang Y, Liu J, Gao C, Qiu J-L (2014) Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery

mildew. Nat Biotechnol **32**: 947–951

Wang Y, Li J (2011) Branching in rice. Current Opinion in Plant Biology **14**: 94–99

Wellman-Labadie O, Zhou Y (2010) Health Policy. Health policy **95**: 216–228

White PJ, Broadley MR (2009) Biofortification of crops with seven mineral elements often lacking in human diets - iron, zinc, copper, calcium, magnesium, selenium and iodine. New Phytologist **182**: 49–84

Wolfe M (1985) The current status and prospects of multiline cultivars and variety mixtures for disease resistance. Annu Rev Phytopathol **23**: 251–273

Xu K, Xu X, Fukao T, Canlas P, Maghirang-Rodriguez R, Heuer S, Ismail AM, Bailey-Serres J, Ronald PC, Mackill DJ (2006) Sub1A is an ethylene-response-factor-like gene that confers submergence tolerance to rice. Nature **442**: 705–708

Xu X, Liu X, Ge S, Jensen JD, Hu F, Li X, Dong Y, Gutenkunst RN, Fang L, Huang L, et al (2011) Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. Nat Biotechnol **30**: 105–111

Yamaki S, Miyabayashi T, Eiguchi M, Kitano H, Nonomura K-I, Kurata N (2010) Diversity of panicle branching patterns in wild relatives of rice. Breed Sci **60**: 586–596

Yamamoto E, Yonemaru J-I, Yamamoto T, Yano M (2012) OGRO: The Overview of functionally characterized Genes in Rice online database. Rice **5**: 1–10

Yan J, Kandianis CB, Harjes CE, Bai L, Kim E-H, Yang X, Skinner DJ, Fu Z, Mitchell S, Li Q, et al (2010) Rare genetic variation at Zea mays crtRB1 increases. Nature Publishing Group **42**: 322–327

Yonemaru JI, Ebana K, Yano M (2014) HapRice, an SNP Haplotype Database and a Web Tool for Rice. Plant and Cell Physiology **55**: e9–e9

Yoshida H, Nagato Y (2011) Flower development in rice. Journal of Experimental Botany **62**: 4719–4730

Yu J, Hu S, Wang J, Wong GK-S, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, et al (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. indica). Science **296**: 79–92

Zhang C-Q, Xu Y, Lu Y, Yu H-X, Gu M-H, Liu Q-Q (2011) The WRKY transcription factor OsWRKY78 regulates stem elongation and seed development in rice. Planta **234**: 541–554

Zhang D, Yuan Z (2014) Molecular Control of Grass Inflorescence Development. Annu Rev Plant Biol **65**: 553–578

Zhang Z, Zhao H, Tang J, Li Z, Li Z, Chen D, Lin W (2014) A Proteomic Study on Molecular Mechanism of Poor Grain-Filling of Rice (Oryza sativa L.) Inferior Spikelets. PLoS ONE **9**: e89140

Zhao H, Yao W, Ouyang Y, Yang W, Wang G, Lian X, Xing Y, Chen L, Xie W (2015) RiceVarMap: a comprehensive database of rice genomic variations. Nucleic Acids Research **43**: D1018–D1022

Zhao K, Tung C-W, Eizenga GC, Wright MH, Ali ML, Price AH, Norton GJ, Islam MR, Reynolds A, Mezey J, et al (2011) Genome-wide association mapping reveals a rich genetic architecture of complex traits in Oryza sativa. Nature Communications **2**: 467–10

Zhao K, Wright M, Kimball J, Eizenga G, McClung A, Kovach M, Tyagi W, Ali ML, Tung CW, Reynolds A (2010) Genomic diversity and introgression in O. sativa reveal the impact of domestication and breeding on the rice genome. PLoS ONE **5**: e10780