Phenotypic Analysis and Genetic Dissection of Seed Imbibition and Early Germination in Maize (*Zea Maize* L.)

By

Scott Clarence Stelpflug

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

Plant Breeding and Plant Genetics

at the

UNIVERSITY OF WISCONSIN-MADISON

2015

Date of final oral examination: 12/02/2015

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

Shawn M. Kaeppler, Professor, Agronomy

Natalia de Leon, Associate Professor, Agronomy

Edgar Spalding, Professor, Botany

Patrick Masson, Professor, Genetics

Jeffrey Endelman, Assistant Professor, Horticulture

PHENOTYPIC ANALYSIS AND GENETIC DISSECTION OF SEED IMBIBITION AND EARLY GERMINATION IN MAIZE (*Zea mays L.)*

Under the supervision of Professor Shawn M. Kaeppler

ABSTRACT

The success of seed germination and the establishment of a normal seedling are fundamental in the propagation of plant species grown from seed, having both economic and ecological importance. Imbibition, the first phase of germination, involves a rapid uptake of water by the dry seed, and has been shown to affect germination rate, overall stand count, and seedling vigor, all of which are important pre-requisites ultimately affecting optimum yield potential. The genetic architecture of maize seed traits is incredibly complex due to the lack of correspondence between transcriptional, proteomic, and metabolome levels attributable to dynamic post-transcriptional and post-translational regulation in maize seed tissues. Additionally, maize seed traits are influenced by the effects of seed production environment, seed parent genetic effects due to differential gene dosages in unique seed compartments, and/or parental imprinting of alleles, further adding to their genetic complexity. Hence, the goal of this thesis was to utilize of a systems-based 'omics' approach, integrating image-based phenomics, quantitative genetics, transcriptomics and proteomics to dissect the genetic architecture of seed imbibition and early germination in maize.

To accomplish these goals, we first created and evaluated the efficacy of a semi-automated, high-throughput image-based seed imbibition phenotyping platform across time to

measure initial seed size, imbibition rate *K*, and percent increase in seed surface area 24 hours post-imbibing (%ISSA). Second, we identified sources of variance for these traits due to experimental design, seed production environmental effects, and genetic combining ability. Third, we evaluated the relationships among these seed imbibition traits, as well as their repeatabilities and heritabilities. Subsequently, the genetic architecture of quantitative variation in maize seed imbibition traits was dissected in a diverse maize population using genomic, transcriptomic, and proteomic datasets, identifying putative candidate genes and molecular pathways associated with maize seed imbibition. Finally, we also demonstrated a highly significant relationship of imbibition rate with radicle emergence time and total germination percentage, indicating that our rapid phenotyping platform exhibits potential agronomic utility. In summary, this thesis constitutes a substantial effort towards establishing research and breeding strategies expected to improve early germination characteristics of maize.

ACKNOWLEDGEMENTS

This thesis could not have been accomplished without the help, support, and mentorship from of a number of people involved during different stages of this research. First and foremost, I would like to express my gratitude to Professor Dr. Shawn M. Kaeppler for his advice, guidance, and exorbitant patience exhibited during my years developing and working under his supervision. Not only did Shawn give me the opportunity to work in his lab, but he also encouraged me to think critically and independently, to multi-task and progress on several diverse projects, and to grow and mature as a person. Because of his mentorship, I now feel ready to tackle future challenges in my future career, and more importantly, in life. I would also like to thank my co-advisor Professor Dr. Natalia de Leon for her mentorship throughout this journey and for the initial opportunity to research at UW-Madison as an undergraduate. Next, I would like to thank Nathan Miller for his work creating the initial concept for this project, and for his time developing code for the image-analysis pipeline. Without his and Dr. Spalding's lab and time, this project would not have been completed. I would also like to acknowledge the other co-members of my committee, Drs. Edgar Spalding, Patrick Masson, and Jeffrey Endelman for their advice and time devoted to my thesis.

My gratitude also extends to Dustin Eilert, Marina Runge, Wilson Craine, Jimmy Flannery, and other former and current graduate and undergraduate students of the field corn research group. Our research group truly operates as a team, and contributions from everyone were necessary in order to achieve my goals. Specifically, I would like to thank Dr.'s German Muttoni and James Johnson for taking me under their wing towards the beginning of my time as a graduate student. They helped teach me basic fundamentals of plant breeding and also helped teach me how to be a professional during graduate school. I'd also like to thank my office mates

Nick Haase, Calli Anibas, Joe Gage, and Rajan Sekhon for their intellectually stimulating conversations. I will always cherish the memories and laughs. Furthermore, I am deeply grateful for the rest of the Plant Breeding and Plant Genetics faculty, the Agronomy department staff, and the entire graduate student body for making my time at UW-Madison extremely enjoyable and personally enriching.

I would also like to acknowledge the support provided by my family and friends throughout this process. My family has always supported and encouraged me to pursue my dreams. Specially, I want to thank my parents; their love and moral values have helped shape me into the person I am today- my mother, Paulette, for her sacrifice of staying home with me and my brothers growing up and for fostering my quest for scientific knowledge from a young age. I'd also like to thank my father, Dick, a fellow corn breeder whose dedication, love, and passion for his job spurred me to get involved in the field of plant breeding. Additionally, I am grateful for my two little brothers, Mike and Alec, for always being there no matter what, and to my wonderful girlfriend, Lea, for her unconditional love and support throughout this endeavor. Lastly, she also helped me plate a lot of kernels when I was busy!

TABLE OF CONTENTS

LIST OF TABLES

Chapter 2

Chapter 3

LIST OF FIGURES

Chapter 1

Chapter 2

Chapter 3

LIST OF APPENDICES

Chapter 2

Chapter 3

CHAPTER 1. AN OVERVIEW OF THE FACTORS GOVERNING MAIZE KERNEL IMBIBITION (LITERATURE REVIEW)

1.1 Maize Kernel Anatomy and Composition

Formation and structure of the maize kernel

Grass plants produce one-seeded dry fruits known as caryopses (popularly known as grain). These types of fruits are monocarpelate and indehiscent, having the ovary wall tightly bound to the seed coat. In maize (*Zea mays L.*), the caryopsis is also known as the kernel. A mature maize kernel is composed of three main parts: embryo, endosperm and carpel wall (seed coat and pericarp) (**Figure 1**).

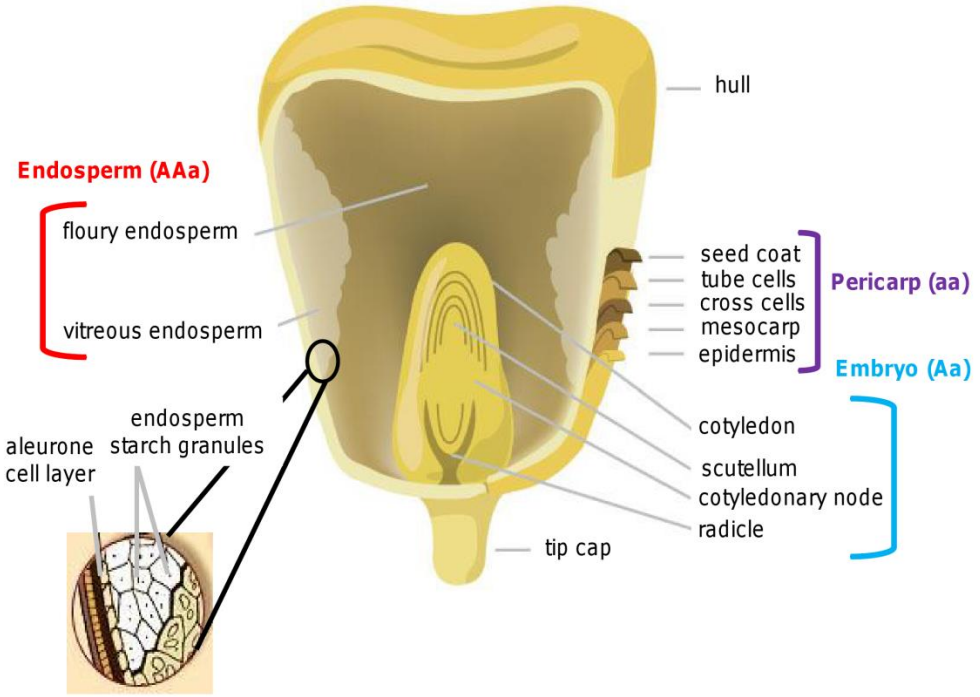


Figure 1. **The anatomy and structure of the maize kernel.** Figure was adapted from Britannica Encyclopedia, 1996.

Initiation of maize kernel development first requires fertilization of the egg and polar nuclei by sperm cells from mature maize pollen grains (KIESSELBACH 1999). Flowering of the monoecious, diclinous maize plant occurs when it has completed the vegetative (VT) stage of development. Pollen grains are formed in the staminate flowers in the tassel. Mature pollen grains are tri-cellular, with two generative sperm cells and one vegetative cell, which forms the pollen tube. Silks are formed from the pistillate flowers which develop from the ear shoot meristem, and their emergence signifies the onset of the first reproductive (R1) stage (RITCHIE *et al.* 1993). Upon contact with silk, pollen grains germinate and pollen tubes develop. The pollen tube grows down the length of an individual silk and enters the embryo sac to release the two sperm nuclei (KIESSELBACH 1999). One of the sperm nuclei fuses with the two polar nuclei of the embryo sac to form triploid endosperm tissue, while the other sperm nucleus fuses with the egg to form a diploid embryo (KIESSELBACH 1999). Upon fertilization, the kernel commences development, and after roughly 55-65 days, the mature maize kernel is formed (TRUE 1893).

Endosperm

As mentioned previously, the endosperm is the product of the second fertilization event during double fertilization (KIESSELBACH 1999). The endosperm is the largest portion of the kernel by dry weight (80-84% and is economically the most valuable portion of the kernel (WOLF *et al.* 1952). At maturity, the endosperm is composed of approximately 90% starch, 7% protein, and smaller quantities of sugars, oils, minerals, and other compounds (WATSON *et al.* 2003). The endosperm is composed of three major cell types: the starchy endosperm, the basal transfer layer, and the aleurone layer (OLSEN 2001).

Endosperm transfer cell layer

The transfer cell layer is located at the basal part of the grain and functions in nutrient uptake from the mother plant during seed development. There are no vascular connections between the maternal plant and developing caryopses, and all nutrients entering the seed pass through a specialized group of cells called transfer or basal endosperm cells by symplastic or apoplastic methods (HUEROS *et al.* 1995; HUEROS *et al.* 1999). Moreover, endosperm transfer cells are characterized by the occurrence of cell wall ingrowths, which increase the surface of the cellular membrane up to 22-fold and make endosperm transfer cells proficient in the uptake of nutrients from adjacent maternal vascular tissues to the endosperm. Sucrose synthase and the cytoskeleton also probably play a primary role in the wall ingrowth formation (WANG *et al.* 1994). In general, this region of the endosperm plays a large role in solute transport of amino acids, sucrose, and monosaccharides, which is consistent with the expression of these transporters and transport-associated genes in this tissue (THOMPSON *et al.* 2001; ZHENG AND WANG 2010).

Starchy endosperm

The starchy endosperm comprises the major part of the endosperm and is formed by relatively uniform cells in shape and context. Mature starchy endosperm is formed by dead cells filled with starch granules and protein bodies. These reserves, which account for most of the mass of dead starchy endosperm cells, are not broken down until germination has been triggered and represent the basic fuel for the growing seedling until autotrophy (KEELING AND MYERS 2010). Starch exists in the form of granules, as a mixture of two polymers of glucose units, amylose and amylopectin, in a ratio of 1:3, respectively (BEMILLER 2007). Amylose in the maize

kernel exists in a linear chain-like structure, while amylopectin is a large branched polymer (HEDLEY 2001). Starch synthesis utilizes glucose residues originating from the degradation of sucrose (ZAMSKI AND SCHAFFER 1996). The synthesis process begins in the central crown of the endosperm and proceeds globally in a radial fashion toward the kernel base (BOYER AND SHANNON 1986). The nutrients are made available as a result of hydrolysis by enzymes produced in the embryo and in the aleurone layer. The major protein storage structure in maize endosperm, termed protein bodies, are synthesized in the endoplasmic reticulum of the endosperm and are comprised almost entirely of alcohol-soluble prolamins (zeins) and glutelins; these are membrane bound structures which have polyribosomes associated with their outer surface (LARKINS AND HURKMAN 1978). Maize prolamins (zeins) are divided into different types (α-, β-, γ-, and δ-zeins) that differ in their amino acid composition and structural properties (HERMAN AND LARKINS 1999; SHEWRY AND HALFORD 2002). Zeins and glutelins represent approximately 60% and 30%, respectively, of the total protein in typical maize endosperm (LEE *et al.* 1976).

Although oils represent a minor proportion of the main reserves accumulated in the starchy endosperm, triacylglyceride deposits are present in small quantities throughout the starchy endosperm (HARGIN *et al.* 1980; BARTHOLE *et al.* 2012). Oil bodies in the starchy endosperm have diffuse boundaries and are fused both with each other and with protein vacuoles during grain development, forming a continuous oil matrix between the protein and starch components (HENEEN *et al.* 2008). Starchy endosperm cells undergo cell death during seed maturation. Unlike programmed cell death of most other plant cells, including aleurone cells, the death of starchy endosperm cells is not followed by a rapid destruction of the corpse (YOUNG AND GALLIE 1999).

Aleurone layer

The maize aleurone layer consists of a single layer of cells having a regular, cube-like shape with thick walls and prominent nuclei. The cytoplasm of mature aleurone cells is filled with organelles, the most prominent of which are the protein storage vacuoles (PSVs) (BETHKE *et al.* 1998). The also contain oleosomes embedded in their membranes (BETHKE *et al.* 1998). PSVs store a small amount of non-starch-based carbohydrates (JACOBSEN *et al.* 1971) and abundant mineral reserves that are chelated into phytin (STEWART *et al.* 1988). Mitochondria, ER, glyoxysomes and golgi are also abundant in aleurone cells (JONES 1969). While performing some storage functions, the primary role of the aleurone cell layer is digestive (BETHKE *et al.* 1998). Upon imbibition, the embryo produces gibberellins (GA), which induce cells in the aleurone layer to secrete amylases and proteases that break down the stored starch and proteins in the dead starchy endosperm (FATH *et al.* 2000). The aleurone is also a major site of mineral storage (STEWART *et al.* 1988) and serves to protect the nutrient-rich endosperm by expressing an array of stress and pathogen protective proteins (JERKOVIC *et al.* 2010).

Embryo

The embryo of grass species at grain maturity is composed of two main parts- the embryonic axis and the scutellum. The embryonic axis contains a coleoptile, which is a sheath surrounding five to six preformed embryonic leaves. The scutellum represents the single cotyledon in monocotyledonous embryos and is attached to the embryonic axis at the scutellar node. The embryo has much higher oil (35-40%), lipid-soluble vitamins, and protein (18%) concentrations compared to the rest of the seed and is an important assembly point for many of the compounds essential for germination (MCDONALD 1994; BARTHOLE *et al.* 2012).

Scutellum

Dissected scutellum tissue constitutes 11% of the dry seed weight and about 90% of the embryo (MCDONALD 1994). The scutellum consists of four cell types in total (SWIFT AND O'BRIEN 1970; SWIFT AND O'BRIEN 1972): (1) the epithelium, a single cell layer bordering the starchy endosperm. During germination, epithelial cells secrete hydrolytic enzymes which diffuse into the endosperm where they digest starch and proteins (NOMURA *et al.* 1969). The resulting sugars, peptides and amino acids are then translocated through the scutellum into the embryonic axis (NOMURA *et al.* 1969). The epithelium folds inward in the scutellum, forming "glands" up to 1 μm in length, thus providing additional surface area for enzyme secretion. The walls of the secretory cells are different from other cell walls as they are comprised of only one variety of hemicellulose and very little cellulose; (2) parenchyma, comprising the majority of the scutellum. Parenchymal cells contain dense cytoplasm, some starch granules, and clear lipid-body oleosomes, having thick walls bearing numerous pits and intracellular spaces that facilitate movement of materials among cells; (3) the epidermis, a single layer of cuticulized cells bordering the embryonic cavity; (4) provascular tissue, which differentiates 3-30 hours after imbibition into phloem and xylem in maize. The temporal and spatial differentiation of the scutellar vascular system indicates that it plays an important role in secretive and absorptive functions within the scutellum (SWIFT AND O'BRIEN 1970).

Pericarp

At maturity, the pericarp is tightly bound to the aleurone layer, which completely surrounds the kernel with the exception of the black layer, which represents the compressed remnants of the parental conducting tissue into the seed during its development. The pericarp

makes up 5-6% of the seed dry weight, and is chiefly composed of fiber (87%) and a small component of starch (7%); it plays an important role in regulating water uptake and limiting the entry of fungi or bacteria that may invade a damaged kernel (MCDONALD 1994). It is composed of the following layers of elongated dead cells: epidermis, mesocarp, cross cells, tube cells, and seed coat (WATSON *et al.* 2003). These dead cells become important channels for transporting water when it enters through the tip cap (ECKHOFF AND OKOS 1989). Entry or loss of water through the pericarp is restricted by a waxy coat on the outside of the epidermis (WATSON *et al.* 2003). The layers of the pericarp have different thickness, and the number of rows of cells in each layer may differ. Wolf *et al.* (1952) reported that the epidermis is made up of a single row of cells 0.7-1.0 μm thick, while the mesocarp has several cell rows with at least a tenfold increase in thickness. The overall thickness of pericarp in corn kernels varies from 67-212 μm (WOLF *et al.* 1969). The thickness variation throughout the pericarp is due primarily to differences in compression on the pericarp by other parts of the kernel rather than to differences in the number of cell layers. In maize, the pericarp has the smallest moisture diffusion coefficient of all kernel anatomical components (SYARIEF *et al.* 1984), and removal of the pericarp from the maize kernel causes two orders of magnitude increase in the moisture diffusion coefficient ($\sim 10^{-9}$ vs. $\sim 10^{-7}$) (MUTHUKUMARAPPAN AND GUNASEKARAN 1994). Previous work has established that seed macromolecule composition significantly dictates imbibition and water uptake, in the order of protein > carbohydrate > lipid content (MCDONALD 1994) (reviewed in more detail in section 1.3). Hence, as these characteristics of various seed compartments differ in both genetic origin and structure and may combine to jointly orchestrate imbibition rate in a highly-regulated fashion.

1.2 Physical, Morphological, and Physiological Aspects of Imbibition

Seed germination begins when dry seeds come in contact with water under favorable conditions. During seed germination, the increase of total water content or fresh weight comprises three phases of water uptake (BEWLEY 1997) (**Figure 2**). Phase I of this process, imbibition, is the uptake of water by the dry seed. Imbibition is an essential process initiating seed germination that moves the seed from a dry, quiescent, dormant organism to a resumption of embryo growth. Consequently, it initiates an orderly transition of increased hydration, enzyme activation, storage product breakdown, cellular metabolism, and a resumption of seedling development occurs. Imbibition occurs due to the absorption of water by cell-wall and protoplasmic macromolecules contained within the seed, such as proteins and polysaccharides, wherein water molecules are held by electrostatic forces such as hydrogen bonds (NOGGLE AND FRITZ 1983). During imbibition, the seed rapidly swells and changes in size and shape (ROBERT *et al.* 2008; PRESTON *et al.* 2009). The movement of water into the seed is due to diffusion and capillary action with water moving from a region of higher water potential (the environment) to the seed, which possesses extremely low water potentials attributed to their osmotic and matric characteristics (VERTUCCI 1989). The uptake of water during imbibition is essential to the initiation of cellular metabolism for at least three reasons: (a) to activate enzymes, (b) to solubilize and transport reactants, and (c) to serve as a reactant itself, especially in the hydrolytic digestion of stored reserves of protein, carbohydrate, and lipids (WOODSTOCK 1988).

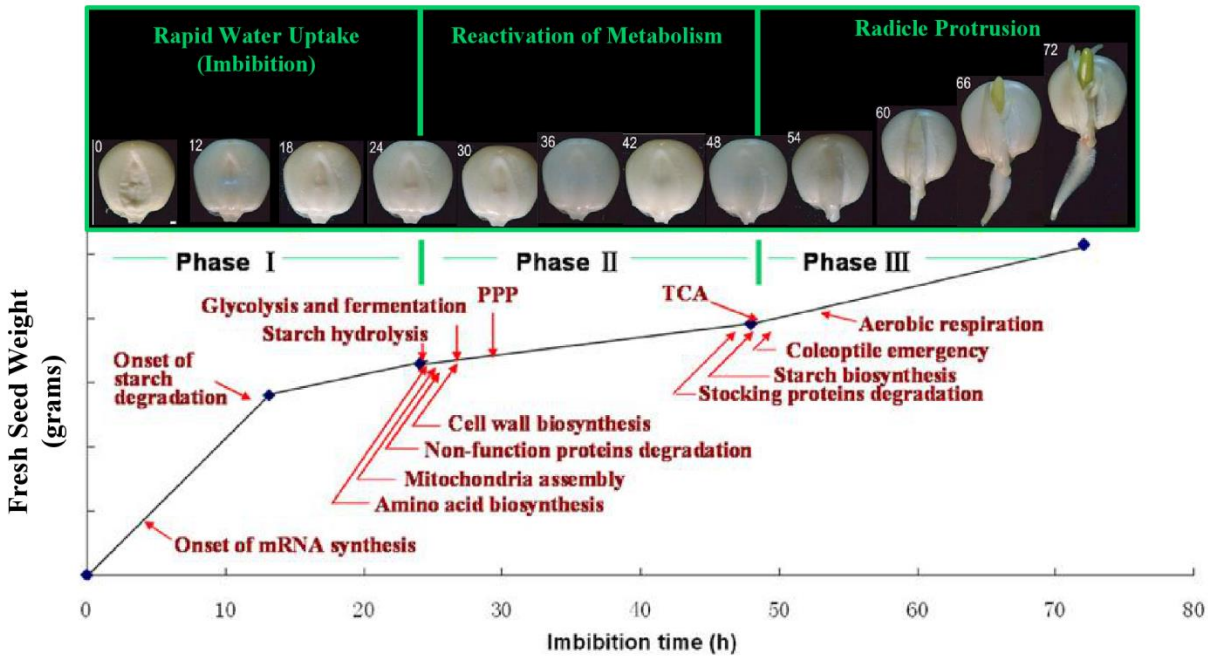


Figure 2. **The triphasic model of water uptake related to seed germination in maize.** Time course images of germinating kernels was produced and adapted from Liu *et al.* 2013, and the graph of physiological processes which occur during germination was adapted from He *et al.* 2009.

Drying and rehydration of a seed from the dry state imposes considerable stress upon the component cells; initial imbibition results in temporary structural perturbations to the seed, particularly to membranes, which lead to an immediate and massive leakage of cellular solutes and metabolites. This leakage is more evident as seeds are deteriorated by age (DUKE *et al.* 1983; SCHOETTLE AND LEOPOLD 1984). While leakage can hasten germination by lowering inhibitor concentrations within seeds (MATILLA *et al.* 2005), it is also a sign of damage to membranes and cellular compartments caused by fast and/or non-homogeneous rehydration (POWELL AND MATTHEWS 1978). This is symptomatic of a transition of the membrane phospholipid

components from a "gel phase" achieved during maturation and drying to the normal, hydrated liquid-crystalline state (CROWE AND CROWE 1992). Similar phenomena can be observed in resurrection plants and pollen that rapidly return from a dry quiescent state to a fully hydrated state (HOEKSTRA *et al.* 1999). In order to cope with the damage imposed during dehydration, storage, and most significantly, rehydration, seeds activate a number of repair mechanisms during imbibition. This includes the repair of membranes, proteins, and damaged genomic DNA, which results from the cumulative effects of temperature, moisture, oxygen, and ROS levels (BRAY AND WEST 2005). The accumulation of chromosomal damage and/or an inability to repair such damage during the imbibition period appear to be significant factors contributing to loss of seed viability during storage (WEITBRECHT *et al.* 2011).

Upon imbibition, the quiescent, dry seed rapidly restarts metabolic activity (**Figure 2**). The structures and enzymes necessary for this initial resumption of metabolic activity are generally present within the dry seed, having partially survived the dry-down phase which terminates seed maturation. Reintroduction of water during imbibition is satisfactory for metabolic activities to resume, with turnover or replacement of cellular components occurring over several hours as full metabolic status is achieved. One of the first changes to the seed upon imbibition is the renewal of respiratory activity, which can be detected within minutes. After a sharp initial increase in oxygen consumption, the rate declines until the radicle penetrates the endosperm and seed coat. The glycolytic and oxidative pentose phosphate pathways (PPP) both resume during imbibition, and the Kreb's cycle enzymes become activated (NICOLAS AND ALDASORO 1979; SALON *et al.* 1988). Germinating seeds of many species frequently produce ethanol (MOROHASHI AND SHIMOKORIYAMA 1972). This is often the result of an internal

deficiency in oxygen that is caused by restrictions to gaseous diffusion by both the internal and surrounding structures of the seed.

Tissues of the mature dry seed contain mitochondria; although these organelles are poorly differentiated as a consequence of maturation drying, they contain sufficient Kreb's cycle enzymes and terminal oxidases to provide adequate amounts of ATP to support metabolism for several hours after imbibition (EHRENSHAFT AND BRAMBL 1990; ATTUCCI *et al.* 1991). Maize kernels repair and activate pre-existing organelles as opposed to making new ones (MOROHASHI AND BEWLEY 1980). The biogenesis of mitochondria in germinating maize embryos involves the synthesis of cytochrome c oxidase subunits encoded by the organellar genome, which is followed within hours by the synthesis of nuclear-encoded subunits (EHRENSHAFT AND BRAMBL 1990). This observation also implies that the coordinated regulation of mitochondrial and nuclear genomes in plants begins during the early stages of germination.

As germination continues post-imbibition, metabolism of the seed becomes engaged in those processes vital for the emergence of the radicle. The process of germination involves a complex "give and take" feedback system between the expansion of the embryo and radicle and the mechanical restraint imposed by the endosperm and testa. In all seeds in which the embryo must penetrate the surrounding structures, embryo growth potential must exceed the mechanical resistance of these tissues. Embryo cells elongate prior to the completion of seed germination of *Arabidopsis*, *Brassica*, *Medicago*, and other species; cell division is not evident in the embryos of these seeds during germination (BARRÔCO *et al.* 2005; GIMENO-GILLES *et al.* 2009; SLIWINSKA *et al.* 2009). Interestingly, a water uptake kinetics study of imbibing Mungbean seeds (*Vigna radiata)* found that seeds incubated with cyclohexamide, a translational inhibitor, during imbibition had decreased water uptake of the embryonic axis, but not of the cotyledons

(CHAKRABORTY AND KAR 2008). This result suggests that imbibition of the cotyledons and endosperm is most likely a passive, physical process, while swelling of the embryonic axis, which is a pre-requisite for turgor driven initiation of axis growth which completes germination, may be an active process mediated by proteins during imbibition. After the initial swelling is completed, all changes in seed size and shape during germination are caused by cell expansion. Expanding plant cells may adjust the extensibility of their cell walls by remodeling the major components of the cell wall, the cellulose microfibrils and/or the pectin/hemicellulose matrix, to accommodate the expanding embryo and radicle. There are many such studies on endosperm weakening by cell-wall remodeling enzymes (CWREs), such as xyloglucan endotransglycosyslase/hydrolase (XTH) (CHEN *et al.* 2002), β-glucanases (LEUBNER-METZGER *et al.* 1995; LEUBNER-METZGER 2002), endo-β-mannases (GROOT AND KARSSEN 1987; GROOT *et al.* 1988), or by other cell wall proteins such as expansins (EXP) (CHEN *et al.* 2001). It is also possible that cell walls are chemically modified by peroxidases and hydroxyl radicles (•OH) *in vivo*, as occurs in germinating cress (PASSARDI *et al.* 2004; MÜLLER *et al.* 2009). Loosening of the wall allows water influx, which drives cell expansion and generates turgor pressure (SCHOPFER 2006). This led to the model that embryo growth during germination depends primarily on changes in cell wall extensibility. These changes are accompanied by progressing vacuolation during late phase II of water uptake. Embryo and endosperm cells are not fully vacuolated during early phase II, but display many small vacuoles (BETHKE *et al.* 2007). In *Arabidopsis*, it is not the radicle itself that undergoes elongation in order to pass through the surrounding structures of the endosperm and testa. Rather, it is a region immediately behind this, the lower hypocotyl and transition zone between the lower hypocotyls and the radicle (SLIWINSKA *et al.* 2009).

Imbibition is not merely an uncontrolled physical event; it is now recognized that chemical conformational events, pericarp effects, and seed quality factors dictate the directed flow of water into the seed (LEOPOLD *et al.* 1989). Thus, any consideration of seed germination physiology and its resultant impact on stand establishment should initially focus on water uptake. The rate and extent to which water imbibition occurs is dependent on four primary factors: (1) the composition of the seed, (2) seed coat permeability, (3) water availability, and (4) temperature (WOODSTOCK 1988).

The effects of seed composition on imbibition

Dry seeds typically possess very low water potentials attributed to their osmotic and matric characteristics, which causes a rapid water influx. These potentials may be as low as -400 MPa in the dry maize kernel (SHAYKEWICH 1973). The low water potentials are a consequence of the relationship of water with components of the maize kernel, which have discrete anatomical regions (pericarp, endosperm, and embryo) containing varying proportions of the three major classes of macromolecules found in seeds: carbohydrates (mostly in the form of starch), proteins, and lipids/oils (previously described above in chapter 1.1). Maize dent whole kernels are, on average, 70% starch, 2% sugars, 14% proteins, 4% oils, and 9% fiber. However, maize kernels exhibit substantial genetic and phenotypic variation for the relative abundance and types of these macromolecules. For example, a phenotypic analysis of kernel composition (given by % dry matter) of 5000 recombinant inbred lines (RILs) from the maize nested association mapping (NAM) population showed dynamic ranges of starch (59.7-73%), protein (10.8-17.7%), and oils (2.8-6.4%) (COOK *et al.* 2012). Highly significant ($p < 0.0001$) phenotypic correlations were

detected between starch and protein (r = -0.66), starch and oil (r = -0.41), and protein and oil (r = 0.32). From this same study, the authors conducted a joint, stepwise regression linkage analysis of NAM and identified 21 starch, 26 protein, and 22 oil QTLs, which collectively explained 59%, 61% and 70% of the total variation, respectively; this indicates that many genes of small effect dictate kernel composition. Additionally, the complex genetic architecture of starch, protein, and oil content has been demonstrated in the inbred line (IL) long-term selection experiment, in which more than 100 generations of recurrent selection has increased oil and protein content to approximately 20% and 27%, respectively (MOOSE *et al.* 2004). The continuation of phenotypic gain from selection of kernel composition provides convincing evidence that these traits are controlled by many genes, which is confirmed by the Cook *et al.* (2012) study.

Importantly, these macromolecules accumulate to drastically dissimilar proportions in different anatomical regions of the maize kernel: the majority of the starch is contained within the endosperm (88% dry weight), the majority of proteins (18% dry weight) and lipids (30-35%) are contained within the embryo, and the majority of fiber is found in the pericarp (87% dry weight). The positive correlation of seed lipid and protein content reported by Cook *et al.* (2012) most likely is due to the fact that the abundance of both of these macromolecules is in the embryo, while the strong negative correlations of lipid and protein with starch may indicate that there is some metabolic trade-off between increasing embryo size (and hence lipids and proteins) with endosperm content (starch).

It has been well established that seed macromolecule composition significantly dictates imbibition and water uptake, in the order of protein > carbohydrate > lipid content (MCDONALD 1994). Thus, seeds containing higher protein content will imbibe more water than seeds

containing comparatively more starch or oil. Proteins are zwitter-ions that exhibit both negative and positive charges that attract the highly charged polar water molecules. Primary water sorption sites on protein molecules have different hydrophilicity. Previous work has calculated that the highest number of H_20 molecules per mole specific site was for $–NH_2$ and –COOH groups, a medium value was found for the phenyl-OH groups, and the lowest hydrophilicity was exhibited by aliphatic –OH, -CONH-, $-CONH_2$ and –NH- groups (LEEDER AND WATT 1974). Therefore, the water sorption power of grains depends not only on protein content but also on protein composition (RATKOVIC AND PISSIS 1997). In contrast, starches such as amylose have little affinity for water while lipids that have no charges have no affinity for water. For example, starch from the maize kernel endosperm is far more hydrophobic than the embryo, as can be demonstrated by greater water contents at a given relative humidity (CHUNG AND PFOST 1967) and water absorption by the embryo is ten times faster than the endosperm (STILES 1948). Furthermore, previous work in soybean showed that the embryonic axis hydrated more rapidly and completely than any other part of the seed due to its higher protein content (MCDONALD *et al.* 1988b). Similar observations have been reported in five legume species where the embryonic axis characteristically had higher water-binding properties than the cotyledons (VERTUCCI AND LEOPOLD 1987). Not surprisingly, seeds of different species imbibe water differently. Ratkovic and Pissis (1997) examined hydration capacity ("a" = weight of water absorbed (g) / weight of seed dry matter (g)) of 11 different species, which ranged from a = 0.44-0.50 for small grains and cereals (wheat, triticale, and maize) to a = 1.30-1.80 for large seeded legumes (white lupin, chickpea, pea). This large variation in water uptake could largely be explained by differences in starch and protein amongst species, with protein content being highly positively correlated to water uptake (r = 0.84) (RATKOVIC AND PISSIS 1997). Given the significant effect of seed

composition on imbibition, and the substantial, natural genetic variation which exists within maize, it is expected that natural genetic variation in seed composition will significantly affect kernel imbibition rates in maize.

The effects of seed coat morphology and permeability on imbibition

The preceding discussion considered the effect of seed composition on imbibition. The entry of water into seeds is, however, first dependent on the permeability and morphology of the seed coat or the fruit coat of the seed (MAYER AND POLJAKOFF-MAYBER 1982). Many species, not including maize, possess seeds with hard seed coats (impermeable to water) that have small elongated pores and a high density of waxy material embedded in the testa epidermis; this impermeability is frequently the cause of dormancy in several species, as it prevents the seed from swelling even in favorable conditions (CALERO *et al.* 1981; YAKLICH *et al.* 1986; WATSON *et al.* 2003).

In maize, the outermost layer of the kernel is the pericarp, which completely surrounds the caryopsis except for the pedicel end. It is composed of the following layers of elongated dead cells: epidermis, mesocarp, cross cells, tube cells, and seed coat (WATSON *et al.* 2003). These dead cells become important channels for transporting water when it enters through the tip cap (ECKHOFF AND OKOS 1989). Frequently, the seed coat exhibits selective permeability toward certain substances. In soybean, the seed coat is extremely hydrophilic and is able to absorb as much as four times (approximately) its fresh weight in water (MCDONALD *et al.* 1988a). In maize, the pericarp also has the smallest moisture diffusion coefficient of all kernel anatomical components (SYARIEF *et al.* 1984), and removal of the pericarp from the maize kernel causes two orders of magnitude increase in the moisture diffusion coefficient ($\sim 10^{-9}$ vs. $\sim 10^{-7}$)

(MUTHUKUMARAPPAN AND GUNASEKARAN 1994). This water holding capacity is essential for the seed to avoid imbibitional injury from the rapid inrush of water into tissues with very negative water potential, culminating in damage to the embryo via disruption of membranes and leakage of solutes (POWELL AND MATTHEWS 1978; TULLY *et al.* 1981; DUKE *et al.* 1983). This appears to be the case for pea seed whose testa has been removed (POWELL AND MATTHEWS 1978). Any damage to the integrity of the pericarp, or its complete removal, can influence the rate of water uptake (WATSON AND SANDERS 1961; RUAN *et al.* 1992; MUTHUKUMARAPPAN AND GUNASEKARAN 1994; RAMOS *et al.* 2004), increase the incidence of imbibitional chilling injury (CAL AND OBENDORF 1972; TULLY *et al.* 1981), and decrease field emergence (MEYERS 1924; TATUM AND ZUBER 1943).

While it would be expected that the process of water uptake is uniform throughout the seed, studies examining this process in maize have demonstrated that imbibition is not uniform. The maize seed first uptakes water through the tip cap, with capillary forces rapidly moving the water both tangentially and radially through the cross and tube cells of the pericarp to the top of the kernel (WOLF *et al.* 1952); these cells may increase the water storage capacity around the radicle tip ensuring a ready source of water for turgor pressure essential for germination. Microscopic nuclear magnetic resonance imaging (MRI) of corn kernels absorbing water demonstrated that water subsequently diffuses much more rapidly into the embryo relative to the endosperm (RUAN *et al.* 1992) (**Figure 3A**). A slower, more progressive wetting front simultaneously moves through the pericarp and into the endosperm that is not complete even after 48 hours of imbibition (MCDONALD 1994). A study of yellow and white corn hybrids revealed that water absorption rate was highly correlated to the top ($r = 0.60$, $p < 0.0001$) and middle abgerminal side ($r = 0.41$, $p < 0.0001$) thickness of the pericarp, and to the following

physical properties: major diameter (r = 0.31, p < 0.0001), volume (r =0.27, p < 0.05), face area (r = 0.31, p < 0.05), kernel density (r = -0.26, p < 0.05) and bulk density (r = -0.30, p < 0.05) (RAMOS 2009). The continuous formation of cracks inside kernels during the sorption process is also thought to increase the volume of the grain (GROSH AND MILNER 1959; FAN *et al.* 1962), and a subsequent scanning electron microscopy study related to maize milling water absorption rates revealed that the internal porosity of the kernel may have influenced the water absorption rate (RAMOS 2009) (**Figure 3B**).

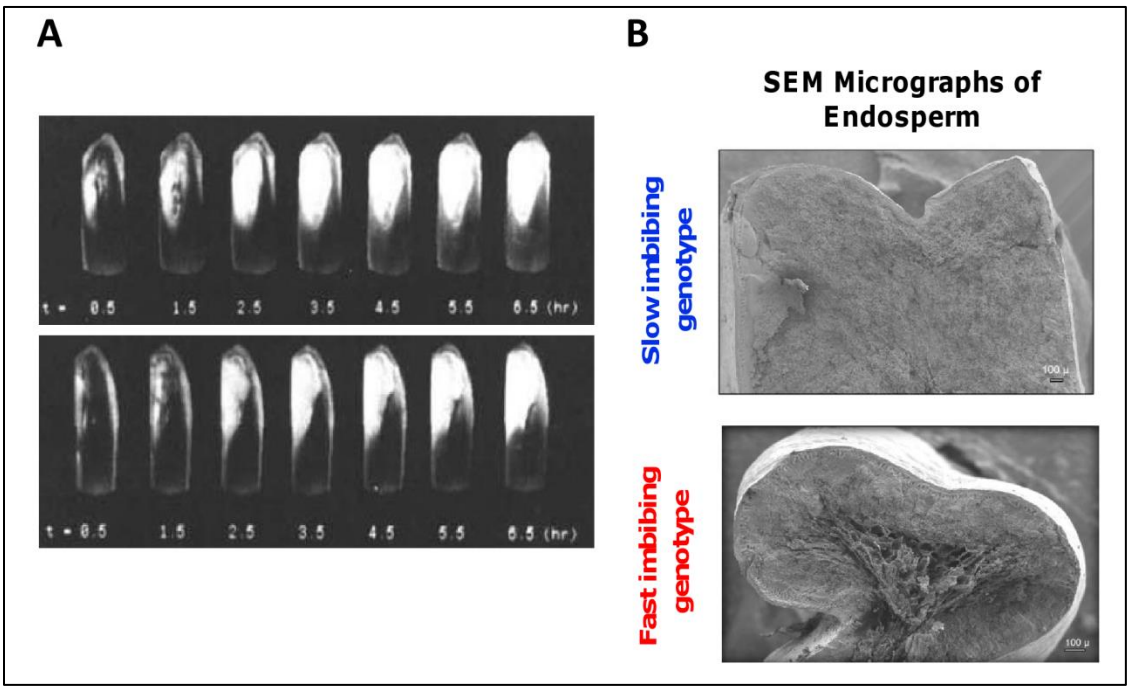


Figure 3. Water uptake by maize kernels is not homogenous and is differentially regulated by unique seed compartments. (A) MRI imaging of water uptake during a time course of imbibing maize seeds (water intensity shown by white coloration) (from Ruan *et al.*, 1992). (B) SEM imaging of maize seeds showing differential imbibition rates indicates that kernel porosity and air spaces influences water uptake rate (imaged adapted from Ramos *et al.*, 2009).

Additionally, various external factors can cause changes in the permeability of the seed coat or pericarp. For example, heat-killed maize kernels often imbibe water more rapidly than the corresponding viable kernels, either due to an increase in permeability of the seed coat or an alteration of membrane organization (BLACKLOW 1972). Previous work also showed that the seed coats of various species showed different permeabilities to water, and was able to relate these differences to the composition of the seed coat and especially to hydrophobic lipoid, tannin, and pectin components (DENNY 1917). Further corroborating this observation, in a comparison between permeable and non-permeable soybean cultivars, (SHAO *et al.* 2007) found that the cuticle of the impermeable cultivar contained a disproportionately high amount of hydroxylated fatty acids relative to the permeable ones, and that a brief treatment with hot alkali solution released the hydroxylated fatty acids, creating holes in the cuticle which caused the seeds to become more permeable. The effect of testa morphology and permeability on germination has also been shown in additional species; *Brassica* seeds with different testa morphology show altered germination characteristics (ZENG *et al.* 2004; MATILLA *et al.* 2005), and *Arabidopsis* testa mutant seeds with reduced pigmentation are more permeable to tetrazolium salts than the wild type with seeds showing a lower dormancy and differing in hormone sensitivities during germination (DEBEAUJON AND KOORNNEEF 2000; NORTH *et al.* 2010).

While *Arabidopsis* seed coat development, including the genes and hormones involved in this process have been studied in detail, little is known about the changes of the seed coat's mechanical and biochemical properties that ultimately lead to testa rupture (HAUGHN AND CHAUDHURY 2005). Progression of tobacco testa rupture starts near the micropylar seed end that covers the radicle and spreads along the ridges on the testa (LEUBNER-METZGER 2002).

Progression of tobacco testa rupture is facilitated by channel-like structures underlying the ridges, suggesting pre-determined breaking points. *Arabidopsis* testa rupture also starts at the micropylar end, but it is unknown if pre-determined breaking points exist (WEITBRECHT *et al.* 2011). Future experiments regarding a potential enzyme weakening, spatial water redistribution mechanism in connection with cell elongation and transcriptomic activity before testa rupture are required as they might shed light on how testa rupture is controlled.

The conditions under which seeds are drying have also been shown to affect the permeability of seed coats (MARBACH AND MAYER 1974). When seeds of hard-seeded legumes were dried in the absence of oxygen, the seed coats were water permeable, while in the presence of oxygen, impermeable coats were obtained. The effect of drying in the case of *P. elatius* could be related to the presence of catechol oxidase in the seed coat, which in the presence of air or oxygen led to the oxidation of phenolic compounds, which rendered the seeds impermeable to water. In maize, it was reported that drying of kernels in the field was more closely related to pericarp permeability than to pericarp thickness (PURDY AND CRANE 1967), although thickness was negatively correlated with drying rate during thin-layer, high temperature drying of shelled corn (STROSHINE *et al.* 1987). Such effects could also contribute to part of the parental effects observed on seed behavior.

Additionally, the presence of mucilages in the seed coat in certain species improves the ability to imbibe water and the artificial addition of mucilages has basically the same effect. Mucilage reduces the susceptibility of the seed to soil water tension (HARPER AND BENTON 1966). The ability of seeds to absorb water from the soil as compared to water uptake from solution is determined not only by the osmotic potential of the soil solution, but also by the matric potential of the soil and the contact of the seed surface with soil particles (MAYER AND

POLJAKOFF-MAYBER 1982). Seed coat permeability is an important factor in regulating imbibition rate and chilling injury. For instance, dark-seeded snap bean lines imbibe more slowly than white-seeded lines due to greater seed coat weight and thickness (WYATT 1977). Imbibitional injury was observed in white-seeded bean cultivar at both 20 and 4ºC, while the dark-seeded cultivar was only injured at 4ºC (POWELL 1986). Hard-seededness may be a valuable characteristic for improvement of overall seed quality. Seed viability on the mother plant and after accelerated aging was maintained better in a hard-seeded soybean line than in a cultivar with a permeable seed coat (POTTS *et al.* 1978). However, a coat-imposed dormancy mechanism would also greatly decrease the imbibition rate. For instance, the semi-hard seed (SHS) characteristic has been selected for and evaluated in a snap bean breeding program (DICKSON AND BOETTGER 1982). Semi-hard seeds are those that do not imbibe water during a 24-h period due to having more phenols in the osteosclereid cells and more pectic substances in the palisade cells of the chalazal testa than a cultivar having a permeable seed coat.

The influence of the semi-hard seed trait on plant performance was studied in both laboratory and field trials, with imbibition profiles determined by recording water weight increases overtime to calculate "*K*," the hydration rate. Four breeding lines of snap pea were selected to obtain a range of hydration rates, with one hard-seeded genotype which was most impermeable (NY5345), two intermediate (NY8757 and NY8741), and a very permeable soft-seeded genotype NY8751 (TAYLOR *et al.* 1992). While all genotypes had >85% germination using a standard germination test, the most permeable line was most susceptible to imbibitional chilling injury. However, for the chilled germination test, the most impermeable line with the highest hydration rate exhibited a reduction in germination compared with the other treatments (TAYLOR *et al.* 1992). Field studies which evaluated seedling establishment under both wet and

dry conditions revealed that an increase in the hydration rate *K* was accompanied by a decrease in field emergence (r = 0.98**) (TAYLOR *et al.* 1992). However, the most impermeable line had the poorest emergence. This difference was due to the impermeable nature of the testa and not to chilling injury. The combined results of this study indicate that imbibition rate may greatly impact germination, field emergence and stand establishment, but that the trait exhibits trade-offs; very fast and slow imbibers may be deleterious, and both ends of the phenotypic distribution are differentially effected by the environment, with fast imbibers being advantageous in dry conditions, and slow imbibers being advantages under wet field conditions. The subsequent discussion within this literature review will focus on the important effect of water availability on seed imbibition.

The effects of water availability on the physical process of imbibition

The physical and environmental forces which dictate the rate of water imbibition by seeds are complex. The ability to imbibe water is dependent on cell water potential, with water moving from a region of higher to lower water potential. Water potential is a relative term, indicated by the Greek letter psi = Ψ, and is measured in megapascals (1 MPa = 10 bars = 10^7 ergs cm^{-3}), with pure water assigned the arbitrary value of zero. The water potential gradient between the seed and its surroundings is thus indicative of the force, and in some cases the speed, with which seeds may uptake water. Water potentials of mature dry maize kernels may be as low as -400 MPa (SHAYKEWICH 1973).

Although the seed water potential has been described as a single entity, it is in fact the result of three distinct forces (WOODSTOCK 1988):

(1) Cell wall matric forces (represented by Ψ_m). Cell walls and intracellular inclusions such as mitochondria, ribosomes are characterized by the presence of membranes. These membranes possess charges that attract water molecules and contribute to the total cell water potential. Hydrophilic proteins within cells may also attract water.

(2) Cell osmotic concentration (represented by Ψ_S). The greater the concentration of soluble compounds, the greater the attraction for water.

(3) Cell turgor pressure (represented by Ψ_P). Turgor pressure increases as cell protoplasts swell and press against the cell wall. Unlike the cell wall matric forces and osmotic concentration that attract water molecules into a cell, turgor pressure, which is the result of a restraining force of cell walls, tends to slow down water uptake.

Of all these factors, it is the matric potential of cell walls, their contents, and proteins which are primarily responsible for imbibition, whereas the osmotic potential becomes more important during germination (WOODSTOCK 1988). It should be noted that although imbibition and osmosis are both examples of water diffusion, osmosis refers to water movement across a differentially permeable membrane, whereas imbibition refers to the taking up of water by a colloidal system which results in swelling (NOGGLE AND FRITZ 1983).

The soils in which seeds are planted also exhibit their own water potentials. The physical properties of soils determine the retention and conductivity of water. For example, it is well known that soils heavy in clays, due to possessing positively charged cations, are able to absorb water more vigorously and retain it longer than those possessing high quantities of sand. In effect, the seed and soil water potential must compete with the soil water potential for imbibition to occur. Initially, the difference between seed (~ -100 MPa) and the soil (-3 MPa) is quite large, thus water flows into the seed. However, as imbibition progresses, this difference is reduced in

the immediate vicinity of the seed. If it were not for the conductive ability of soils, imbibition would be quickly halted, however soils exhibit high hydraulic conductivity which self-replenishes the available water surrounding the seed, ensuring continuous uptake of water by the seed. This is important since seeds are sessile and a continuous flow of water is essential for maximum imbibition and hence successful germination.

Seeds also possess their own initial water potential due to water's binding of macromolecules in the seed which may vary by seed lot or genotype, affecting imbibition rates. As seeds take up water, there is an enormous increase in tissue swelling that is likely due to an unfolding of proteins and association of water with the matric forces of cell walls. Throughout the imbibition process, water transitions from acting as a ligand at low seedling moisture, to a solvent or solution at high seedling moisture (VERTUCCI AND LEOPOLD 1986). The greatest relative change in volume occurs between 4 and 8% moisture. Between 8 and 21% moisture content, seed swelling on an absolute volume basis is greatest, and beyond 21% moisture content, very little increase in seed volume is observed, which explains the hyperbolic pattern of seed swelling curves (VERTUCCI AND LEOPOLD 1986). The macromolecular surface of the seed is fully wetted at 35% moisture, since freezing damage to soybean seeds occurs at this moisture content (VERTUCCI AND LEOPOLD 1986).

The effects of temperature and the physical process of imbibition

Temperature influences the rate of imbibition, with low temperatures resulting in slow imbibition, and vice-versa (MURPHY AND NOLAND 1982). However, the effect of temperature on imbibition is complex. Factors dictating the observed temperature effect include decreasing water viscosity with increased temperatures, and a subsequent increase in the kinetic energy of water, and/or the ease with which seed tissues are wetted (VERTUCCI AND LEOPOLD 1983). The kinetic energy is directly proportional to the absolute temperature, while the molecular velocity varies as the square root of the absolute temperature (MAYER AND POLJAKOFF-MAYBER 1982). Leopold (1983) found that the time required for completion of imbibition in soybeans is essentially doubled at 5ºC compared to 20ºC. Temperature also affects the equilibrium volume of polymers; when fully imbibed soybeans were held for more than 1.5 hrs at various steps of elevated temperature (2-35ºC), a volumetric increase at higher temperatures was observed. The equilibrium volume increases with temperature, but the slope of the temperature response is lesser for the temperature range of 1-15ºC than for the range between 20-35ºC (LEOPOLD 1983). As mentioned previously, imbibition may be considered a special case of osmosis. Hence, any effect of temperature on the structure of the colloid and the dimensions of its intercellular spaces probably affects the rate of imbibition. The final volume obtained by seeds imbibing at low temperatures is greater than that resulting from rapid imbibition at higher temperatures (MAYER AND POLJAKOFF-MAYBER 1982).

The temperature coefficient (Q_{10}), a unitless quantity, represents the factor by which the rate (R) of a reaction increases for every ten degree rise in the temperature (T); the rate (R) may represent any measure of the progress a process or reaction; it has often been used to characterize the effect of temperature on the process of imbibition.

The Q_{10} equation is given below:

$$Q_{10} = \left(\frac{R_2}{R_1}\right)^{\left(\frac{10}{T_2 - T_1}\right)}$$

-$\boldsymbol{R_1}$ is the measured reaction rate at temperature T_1 (where $T_1 < T_2$). Note that R_1 and R_2 must have the same unit.

-$\boldsymbol{R_2}$ is the measured reaction rate at temperature T_2 (where $T_2 > T_1$). Note that R_1 and R_2 must have the same unit.

-$\boldsymbol{T_1}$ is the temperature at which the reaction rate R_1 is measured (where $T_1 < T_2$). The temperature unit must be either the Celsius or the Kelvin, and may not be any other unit, such as Fahrenheit. T_1 and T_2 do not need to be exactly 10 degrees apart.

-$\boldsymbol{T_2}$ is the temperature at which the reaction rate R_2 is measured (where $T_2 > T_1$).

If the rate of the reaction is completely temperature independent, it can be seen from the equation above that the resulting Q_{10} will be 1.0. If the reaction rate increases with increasing temperature, Q_{10} will be greater than one. Thus, the more temperature dependent a process is, the higher will be its Q_{10} value. Q_{10} is ~1 for diffusion of ions and molecules in bulk solutions. For typical chemical reactions, Q_{10} values are ~2. For many biological processes, particularly those that involve large-scale protein conformational changes, Q_{10} values are greater than two. Thus, Q_{10} values may be used to infer mechanistic insight about the physiological process under investigation.

Shull (1920) compared the imbibition of *Xanthium* seeds having a semi-permeable membrane and split peas without such a membrane. He noted that swelling of seeds was essentially similar to that of colloids in isolation. The Q_{10} value of imbibition, in both cases was between 1.5-1.8; Shull concluded that no chemical change was involved in the effect of

temperature on imbibition and that it was not markedly affected by the presence of a semi-permeable membrane. Importantly, in addition to *Xanthium*, rate effects in peas and corn also exhibit Q_{10} values < 2.0, suggesting that imbibition is a physical, not an active process, with no "chemical" changes of the water involved (SHULL 1920; SHULL AND SHULL 1924; MURPHY AND NOLAND 1982).

Various attempts have been made to relate the imbibition of seeds to temperature and seed quality. Blacklow (1972) developed an equation for the imbibition of maize:

$$\frac{dW}{dt} = K(f(t) - W) + b \quad [1]$$

In this equation, *K* is a measure of the permeability of the seeds to water during the exponential phase of water uptake, *W* is the water content of the seeds, *f* (*t*) is the water capacity of the seeds over time, and *b* is a measure of the linear phase of water uptake. Blacklow was able to use this mathematical model to predict the imbibition rate of maize kernels under varying environmental conditions, specifically temperature (BLACKLOW 1972).

Leopold (1983) further refined this equation, and instead of focusing on increasing water weights, he investigated the volumetric increases in seed size during imbibition. Leopold calculated swelling coefficient (*K*), the rate of swelling proportion to total swelling, using an integrated form of Pacheles' equation:

$$Kt = -loge^{\frac{(amax-a)}{amax}} \quad [2]$$

where *a* is the volume at time *t*, a_{max} is the volume at full hydration (24 h), and *K* is the swelling coefficient, distinctive for the material under investigation. This equation can subsequently be solved to isolate the swelling coefficient, *K*:

$$K = ((\ln a - \ln a_0)\ t^{-1}) \quad [3]$$

Leopold (1983) reported that volumetric and weight gains in water are not equivalent, and that across most species, swelling volume outstrips weight gain. Leopold reported that the average swelling coefficient *K* for maize kernels was 0.16, which was larger than wheat and oat, but smaller than any of the legume seeds. The volumetric effect of a given increase in water was greater than the amount of water that entered, indicating that hydration does not involve simple insertion of water between polymer plates and spaces in the seed; rather, there are configurational changes in the seed polymers upon imbibition which cause an exaggeration of volume increases over the amount of water entering. When imbibed soybeans were dried to their original dry weight, they did not recover their original volumes, further illustrating this point (LEOPOLD 1983).

Additionally, earlier studies attempted to apply growth models, such as the Lockhart equation (LOCKHART 1965) to predict seed germination rate based on water availability and imbibition (CARPITA *et al.* 1979; SCHOPFER AND PLACHY 1985; BRADFORD 1986). The Lockhart equation describes a steady-state growth rate as a function of cell turgor (Ψ_p) in excess of a minimum turgor or yield threshold (Y) and cell wall extensibility (m):

$$growth\ rate = m(\Psi_p - Y) \quad [4]$$

While the concepts of a turgor yield threshold and cell-wall extensibility are obviously important for understanding embryo expansion, the germination event is not a steady state. Prior to radicle emergence, the growth rate is essentially zero, so that *m* and *Y* are undefined, while after radicle emergence the equation describes radicle growth, not germination per se. Thus, this growth equation is not well suited for analyzing germination because the initiation of radicle emergence represents a change of state rather than a steady-state growth process, particularly for seeds in which the radicle must emerge through a barrier tissue, as is the case with maize.

An alternative approach was then attempted, applying a population-based hydrotime model to describe germination responses to water potential (GUMMERSON 1986). This model is based upon the timing of the radicle emergence event, not the subsequent growth rate, and can accurately describe germination timing of seed populations at different water potentials (BRADFORD 1995). Instead of the growth rate, the germination rate ($GR_g = 1/t_e$) of a particular seed percentage or fraction g is described by:

$$GR_g = \frac{1}{t_g} = \frac{\Psi - \Psi_b(g)}{\theta_H} \quad [5]$$

Where $\Psi_b(g)$ is the threshold or base Ψ that will just prevent germination of percentage g, and θ_H is the hydrotime constant or the accumulated MPa h required for radicle emergence (BRADFORD 1995). The Ψ_b in eqn. 5 is conceptually analogous to the yield threshold (Y) of the Lockhart equation (eqn. 4), and represents the minimum Ψ at which radicle emergence of percentage g of the population can occur. As Ψ increases above this threshold $\Psi_b$value for a particular seed, the time to radicle emergence decreases in inverse proportion. This hydrotime model assumes that all seeds require the same accumulated MPa h θ_H for radicle emergence, but the threshold $\Psi_b(g)$ varies among individual seeds. Seeds with a higher (more positive) Ψ_b take longer to accumulate the required hydrotime than do seeds having a lower Ψ_b. In most cases, the variation in $\Psi_b(g)$ values is normally distributed among seeds in the population, allowing application of probit analysis to determine the values of $\Psi_b(g)$ and θ_H (BRADFORD 1995).

Importantly, this model fit data for sugar beet, lettuce, and tomato seed germination at various combinations of Ψ quite well (GUMMERSON 1986; BRADFORD 1990; DAHAL AND BRADFORD 1994). This research in lettuce determined that germination rate increases linearly with embryo turgor (Ψ_{embryo}) (BRADFORD 1990). These researchers demonstrated that the endosperm presented little physical resistance to radicle growth at the time of radicle emergence,

but its presence markedly delayed germination. Importantly, it was found that the length of the lag period after imbibition before radicle emergence was primarily related to the time required for weakening of the endosperm and not the generation of additional turgor in the embryo (BRADFORD 1990). Cracks and breaks between endosperm cells were observed just prior to visible radicle growth, with the expanding radicle tip apparently forcing the endosperm tissue apart along these cracks. This indicates that Ψ_{embryo} influences the rate at which endosperm weakening occurs (BRADFORD 1990). Additionally, a previously generated maize seed imbibition model revealed that the proportion of the imbibed seed surface was strongly correlated with the seed imbibition rate and consequently with the germination rate (BRUCKLER 1983; GARDARIN *et al.* 2011). Hence, drastic variation in imbibition rates may hasten the remobilization of seed reserves and subsequent germination rates amongst diverse cultivars, and imbibition rate may serve as a useful, rapid trait proxy of radicle emergence and germination time.

1.3 The Impacts of Environmental and Maternal Effects Related to Seed Germination

From a quantitative genetics standpoint, the natural variation exhibited by maize seeds, which exists due to several interacting factors, is difficult to dissect. First, they consist of tissue representing two generations: the pericarp and seed coat are diploid maternal tissue, while the triploid endosperm and diploid embryo are made up of tissues representing the next generation. Second, seed traits are influenced in at least two environments: while maturing they are attached to the mother plant and experience the maternal environment; subsequently, they are dispersed to the germination environment.

Maternal effects, the phenotypic contributions of the mother to her offspring beyond the equal chromosomal contribution expected from each parent, are anticipated to be significant in seeds. Genetic maternal effects are potentially considerable, because (1) the pericarp is genotypically maternal, (2) the endosperm has a larger maternal (2n) than paternal nuclear contribution (n), (3) the endosperm and embryo cytoplasms (including the mitochondrial and chloroplast DNAs) are maternally inherited, and (4) certain genes are imprinted and silenced, depending on maternal or paternal transmission (WATERS *et al.* 2011; WATERS *et al.* 2013). The endosperm contains enzymes important for germination and is also the source of nutrients for the developing embryo (HARVEY AND OAKS 1974). As a consequence of the differential dosage of male and female genes, the female parent may have a more important role in determining the characteristics of this nutrient source. Therefore, it is not surprising that relatively large maternal effects have been found in seed phenotypes of higher plants (ALEXANDER AND WULFF 1985; ANTONOVICS AND SCHMITT 1986; NIEUWHOF *et al.* 1989; SCHWAEGERLE AND LEVIN 1990; BIERE 1991a; BIERE 1991b). Genetic studies have shown a large maternal effect on seed size and germination percentage in maize (EAGLES AND HARDACRE 1979). Additionally, maternal effects

on mineral composition have been documented for several species, such as seed oil in Lupinus (WILLIAMS AND MCGIBBON 1980), fatty acid content in maize (DE LA ROCHE *et al.* 1971), and protein content in dry beans and soybeans (LELEJI *et al.* 1972; SINGH AND HADLEY 1972). In the case of dry beans and soybeans, these differences did not persist into the F_2 generation, suggesting that these are non-cytoplasmic maternal effects. In addition to influencing the rate or timing of germination, maternal genetic effects can also influence the sensitivity of seeds to environmental conditions; for instance, in maize, a maternal effect was found for sensitivity chilling injury during the imbibition phase of germination (CAL AND OBENDORF 1972).

The final class of maternal effects is phenotypic, resulting from the environment or local conditions of the maternal plant when the seed was maturing. Environmental maternal effects are expected to be large in certain seed traits because of the seeds' attachment to the mother plant during at least part of their development and growth. These influences may occur via structure or physiology. The tissues immediately surrounding the developing embryo and endosperm are all maternal. These tissues, the integuments of the ovule and the wall of the ovary, eventually form accessory seed structures which are important determinants of seed dispersal, dormancy, and germination traits, and variation in these traits can carry over to influence the mature phenotype of an individual. The aforementioned seed traits vary with environmental conditions like temperature (LACEY 1996), photoperiod (GUTTERMAN 1992; MUNIR *et al.* 2001), nutrient availability (PARRISH AND BAZZAZ 1985), or soil moisture conditions (GUTTERMAN 2012). Environmental control of germination acts through the seed coat, the endosperm, nutrient and hormone supply (BIERE 1991a; PLATENKAMP AND SHAW 1993; LACEY *et al.* 1997; BASKIN AND BASKIN 1998).

While a considerable number of quantitative genetic studies related to seed characteristics have been conducted (ROACH AND WULFF 1987), very few have jointly analyzed the effects of maternal environment, maternal genetic effects, and their interactions (PLATENKAMP AND SHAW 1993). Hence, in order to isolate sources of additive genetic variation contributing to seed imbibition and germination characteristics, it remains of utmost importance to first isolate the contributions of the maternal genetic and seed production environmental effects.

1.5 The Genomic and Hormonal Underpinnings of Seed Imbibition and Early Germination

The "reawakening" of cellular metabolic processes upon imbibition in seeds is a combination of old and new signals. The ability of dry seeds to transcribe DNA and manufacture proteins in the "dry state" during prolonged after-ripening remains a somewhat contentious claim. However, transient, low-level transcription and translation has previously been documented in air-dried, low-hydrated seeds (LEUBNER-METZGER 2002). More than 12,000 mRNA species are present in dry seeds of *Arabidopsis* (NAKABAYASHI *et al.* 2005) and barley (SREENIVASULU *et al.* 2008); dry rice seeds contain over 17,000 mRNAs (HOWELL *et al.* 2009). These mRNAs are called stored or residual messages, which result from surviving cell desiccation during seed development (HOWELL *et al.* 2009; SANO *et al.* 2012). In almost all cases where proteome analyses demonstrated a protein to be present in dry seeds, the corresponding transcript was also abundant in the dry seed transcriptome (WEITBRECHT *et al.* 2011). Therefore, the abundance of dry seed transcripts simply reflects translation during seed maturation. This conclusion is further supported by the fact that, in many cases, the most abundant transcripts in dry seeds are rapidly degraded upon imbibition, whereas several transcripts with lower abundance in the dry state accumulate during the early germination phase I of imbibition.

Many of the genes encoding residual mRNAs contain abscibic-acid responsive elements (ABREs) with the core motif ACGT in their promoter motifs, and are typically activated during seed maturation (NAKABAYASHI *et al.* 2005). It is possible that stored mRNAs support *de novo* protein synthesis before transcriptional activity is resumed and new messages become available during early stages of imbibition. In dry seeds (e.g. wheat embryos), all of the components needed for the resumption of protein synthesis are present (MARCUS *et al.* 1966; NAKABAYASHI

et al. 2005). Within minutes of rehydration, ribosomes become recruited into polysomal protein-synthesizing complexes, utilizing extant mRNAs (BEWLEY AND BLACK 1994). Involvement of stored mRNAs during germination is supported by experiments on *Arabidopsis* using transcriptional and translational inhibitors. α-Amanitin, an inhibitor of RNA polymerase II does not inhibit germination, while cycloheximide, a protein synthesis inhibitor, clearly prevents radicle emergence; these observations have led to the controversial contention that *de novo* protein synthesis utilizing stored mRNAs is sufficient to allow for the completion of germination (RAJJOU *et al.* 2004). However, although the transcriptional inhibitor did not block germination, it delayed germination in both *Arabidopsis* and rice seeds (RAJJOU *et al.* 2004; HE *et al.* 2011). It seems transcription inhibition influences seedling establishment more than germination *per se* (HE *et al.* 2011). Although stored mRNAs may be utilized during early imbibition, it is unknown which, if any, are essential for the completion of germination.

A number of studies show there is considerable and active transcription taking place during germination, especially during late Phase I and peaking in Phase II, and it is possible this up-regulation of transcription is required for germination to be completed (PRESTON *et al.* 2009). Germination involves a complex "give and take" feedback loop between the expansive force of the embryo and the mechanical restraining force imposed by the endosperm and testa. In all seeds in which the embryo (comprised of the hypocotyl and radicle (RAD)) must penetrate the surrounding structures (micropylar endosperm, positioned over the RAD tip), as is the case for maize, the embryo growth potential must exceed the mechanical resistance of these tissues for radicle emergence to occur (LIPTAY AND SCHOPFER 1983; BRADFORD 1986). Given this "give and take" model of germination, it is important to dissect the transcriptional patterns of various seed compartments to understand the control of germination timing as well as the corresponding

molecular mechanisms. Dekkers *et al.* 2013 analyzed the *Arabidopsis* seed transcriptome by sampling eleven points along the germination time course (dry seed to 38 HAI), including those that allow an analysis of gene expression changes at key events of germination (testa rupture (TR) and endosperm rupture (ER)), with a focus on the micropylar and chalazal endosperm (MCE) and the RAD.

Construction of gene networks revealed two transcriptional phases; the first phase ran from 1 to 25 HAI (before rupture) and is characterized by large transcriptional changes in both up- and down-regulated genes, with most differential gene expression between the 7HAI and 3HAI time point, with the number of differentially expressed genes decreasing thereafter. Between 1 and 3 HAI, differential gene expression was observed, particularly in the MCE relative to the RAD due to its slower imbibition kinetics compared with the more outward-positioned MCE. Several tissue specific genes were found to be up-regulated in MCE (415) and RAD (546), with 84% of expressed genes shared between both tissues (DEKKERS *et al.* 2013). Over-representation analysis of Phase 1 suggests a large overlap in the functional classes that are reactivated in the MCE and RAD, such as genes related to cell wall function, nucleotide and amino acid metabolism, and protein translation. A major difference between the RAD and MCE was related to transport and energy metabolism (lipid metabolism, glycolysis, TCA, and mitochondrial electron transport that are specifically activated in the MCE from 20 HAI. Additionally, it was found that the majority of genes which are up-regulated during seed maturation were inversely down-regulated during imbibition, but nevertheless, these seed maturation-repressed genes were activated in the first transcriptional phase of imbibition and not the second phase which follows TR (DEKKERS *et al.* 2013).

The second phase of transcriptional activation, which runs from TR to the completion of germination, was marked by resumption of differential gene expression, most notably at TR. Over-representation analysis of gene sets that which were up-regulated within the MCE relative to the RAD along the time course revealed that genes related to secondary metabolism, amino acid metabolism, and protein synthesis are activated transiently, while genes which are constitutively higher are enriched for protein degradation, transport, and stress (DEKKERS *et al.* 2013). During the second phase, the majority of the differentially expressed genes are induced rather than repressed, in contrast to phase I. It was also found that a majority of genes up-regulated in the second phase in the MCE significantly overlapped with touch/mechano-sensing genes. Touch induced signaling resulted in a relatively higher abundance of genes related to GO classes such as cell wall associated, calcium binding, disease resistance, kinases, and transcription factors (DEKKERS *et al.* 2013). Additionally, gene expression associated with jasmonate biosynthesis was activated upon TR in the MCE; this plant hormone was recently shown to be a key regulator of plant morphogenesis, enhanced pest resistance upon touch, and also found to engage in hormonal cross-talk with ABA and GA during early phases of germination (CHEHAB *et al.* 2012; ZHOU *et al.* 2015). In contrast, at the later stages of germination, the RAD was particularly enriched for processes related to DNA, RNA, and proteins. Over-representation analysis suggests that energy metabolism (lipid breakdown, glycolysis, TCA, and mitochondrial electron transport) is activated by 38 HAI (DEKKERS *et al.* 2013). Both the MCE and RAD display tissue specific expression patterns with different biological functions, signifying the importance of their specialization and developmental timing related to breakdown of the seed and subsequent germination and emergence of the plantlet.

Additionally, researchers have produced an integrated transcriptome and metabolome dataset describing time points of early imbibition (0, 1, 3, 12, and 24 HAI) in rice (HOWELL *et al.* 2009). Gene expression analysis found a cascading increase in number of differentially expressed genes from 0-12 HAI, with gene expression leveling off thereafter; 1469 and 1276 transcripts are up- and down-regulated between one and three HAI, while only 59 and four transcripts are up- and down-regulated during the first hour of imbibition; the greatest number of significantly different transcriptional changes were observed between 3 and 12 HAI. PageMan analysis revealed that signal transduction pathways, transcriptional regulation, and receptor kinases were up-regulated at early stages of germination (0-3 HAI), while genes associated with the abscisic acid signal transduction were down-regulated. The majority of significant changes in the transcriptome (from 3-12 HAI) were associated with general up-regulation of transcripts encoding components involved in the following cellular processes: cell wall metabolism, lipid metabolism, nucleotide degradation, amino acid synthesis, carbohydrate metabolism, TCA cycle, jasmonate synthesis, cellular transport, organellar protein synthesis, and aspects of secondary metabolism such as isoprenoid and phenylpropanoid biosynthesis (HOWELL *et al.* 2009). These biological processes and transcriptional trends largely agree with results found in *Arabidopsis*, indicating that the majority of the gene expression networks governing germination between species are fairly conserved. In contrast to the limited transcriptional activity occurring during the first hour of imbibition, 25 of 126 profiled metabolites significantly changed levels over the same duration (HOWELL *et al.* 2009). These metabolites are primarily associated with carbohydrate metabolism, the TCA cycle, and glycolysis, implying there is an immediate increase in the activity of these processes to facilitate early, energy-demanding metabolism (HOWELL *et al.* 2009). The combined metabolome and transcriptome analyses identified a clear

program “switch” between 0-1 HAI and 1-3 HAI, signifying that the complex network of genes dictating germination are largely rewired and directed during the early phase of imbibition.

Hormone Metabolism

Hormones have a profound influence over the regulation of seed germination and dormancy. These processes are tightly regulated by a dynamic balance between synthesis and catabolism of abscisic acid (ABA) and gibberellins (GAs); ABA represses germination and promotes dormancy, whereas GAs breaks dormancy and promotes germination. At the molecular level, the ABA/GA balance is in part determined by the antagonistic control of ABA and GA on each other through their reciprocal regulation of the transcription of their metabolic genes (SEO *et al.* 2006; OH *et al.* 2007; TOH *et al.* 2008).

Under favorable environmental conditions, many of the transcripts encoding enzymes involved in hormone metabolism are produced by *de novo* transcription during germination; many have been identified and characterized in *Arabidopsis*. The GA biosynthesis enzyme GA 3-oxidase catalyzes the conversion of inactive forms of the hormone to active forms (GA_9 to GA_4, or GA_{20} to GA_1), whereas enzyme *GA2ox* is a deactivating enzyme which converts active forms of GA to inactive forms. Seeds of the double-knockout mutants *ga3ox1 ga3ox2* do not germinate without the application of GA, demonstrating the requirement of GA biosynthesis in order for seeds to complete germination (KOORNNEEF AND VAN DER VEEN 1980; MITCHUM *et al.* 2006). An important feature of GA and ABA metabolism is their interaction. For example, ABA down-regulates GA biosynthesis and up-regulates GA deactivation, thus modulating seed ABA content (SEO *et al.* 2006; SEO *et al.* 2009). The significance of the balance between ABA and GA signaling for seed germination is perhaps best demonstrated by studies of ABA response mutants

and GA-deficient mutants. In *Arabidopsis*, ABA-insensitive mutations (e.g., *abi1*, *abi2*, and *abi3*) reduce seed dormancy and allow germination at ABA concentrations that are normally inhibitory to wild-type germination (KOORNNEEF *et al.* 1984), whereas mutants deficient in GAs, as mentioned above, fail to germinate (KOORNNEEF AND VAN DER VEEN 1980). Additionally, GAs promote germination by activating a transcriptional cascade of genes encoding enzymes such as α-amylase (VARNER 1964; GUBLER *et al.* 1995), endo-β-1,3 glucanase (LEUBNER-METZGER *et al.* 1995), β-1,4 mannan endohydrolase (BEWLEY 1997; SÁNCHEZ AND DE MIGUEL 1997), and the extensin-like protein *AtEPR1* (DUBREUCQ *et al.* 2000). These enzymes initiate seed germination by hydrolyzing the endosperm and releasing the inhibitory effects of ABA on embryo growth potential (BEWLEY 1997; KOORNNEEF *et al.* 2002). Responses of seeds to environmental factors such as temperature (YAMAUCHI *et al.* 2004; GRAEBER *et al.* 2014), soil nutrition (MATAKIADIS *et al.* 2009) and smoke (NELSON *et al.* 2009) are also well explained by changes in hormone metabolism (SEO *et al.* 2009), although how they specifically influence the appropriate pathways is still being elucidated.

Quantitative Genetic Studies Related to Seed Imbibition Kinetics

In the current literature, there are no quantitative genetic studies mapping quantitative trait loci (QTL) for water uptake/imbibition rate in maize kernels. However, studies in other plant species, such as *Medicago truncatula* and rice (*Oryza sativa*), have successfully mapped QTL related to imbibition rate in seeds, in the context of other germination-related phenotypes. Dias *et al.* 2011 performed a QTL analysis using composite interval mapping (CIM) with 110 microsatellite markers on a recombinant inbred line (RIL) population of 178 lines at sub- and supra-optimal temperatures at early stages of germination in the legume species *Medicago truncatula.* These researchers measured seed mass, initial seed surface area, imbibition rate,

germination rate and percentage, hypocotyl length and width, and early embryo elongation rate using image analysis. The imbibition rate (IR, $mm^2\ h^{-1}$) was estimated from the increase in seed surface area from 0 to 8 h after sowing, which had previously been shown to be correlated with seed water uptake (DEMILLY *et al.* 2007). Ranges of variation of IR represented 80% of the maximum IR value, exhibiting substantial transgressive variation amongst the RILs (DIAS *et al.* 2011).

Correlation analysis among phenotypic traits revealed a significant negative correlation between initial seed surface area and IR (r = -0.65***) indicating a higher imbibition rate for small seeds (DIAS *et al.* 2011). Interestingly, no significant correlation was found between IR and seed weight, indicating that the surface area contact of the seed with water is more important for this trait than seed density. IR was not found to be significantly correlated to any germination traits. QTL analysis revealed QTL for all traits study. One significant QTL was mapped for IR on chromosome 4, which explained 8% of the phenotypic variance and had a LOD score of 3.1 (DIAS *et al.* 2011). This indicates that IR is in fact heritable and controlled by additive genetic variance, however the lack of additional QTL indicate that the trait is probably highly polygenic and controlled by many genes of small effect. However, the QTL mapped for IR did not co-localize with any other important germination traits, indicating the trait may not be dictating emergence rates in *Medicago*. The average QTL support interval for all traits mapped was quite large due to an insufficient number of markers and RILs, ranging from 4 to 16 cM, containing an average of 700 genes, so any putative candidate genes in these intervals were mostly speculative by the researchers. Interestingly, co-localizing with the QTL for IR was the *Arabidopsis* ortholog of *AtTT12* (*TRANSPARENT TESTA 12*), a gene involved in proanthocyanidin biosynthesis (DEBEAUJON *et al.* 2001). It has been suggested that an increase of flavonoids such as

proanthocyanidins present in the seed coat could slow down imbibition by modulating the polarity and conductivity of the seed coat (KANTAR *et al.* 1996; DEBEAUJON AND KOORNNEEF 2000; DEBEAUJON *et al.* 2001).

Additionally, rice researchers have recently mapped QTL for IR in the context of other rice seed germination traits under salt stress (WANG *et al.* 2011). IR was a trait of interest in this study because previous studies in rice have demonstrated that IR was negatively correlated with seed germination under salt stress, and decreasing IR could improve seed germination under these conditions (WANG *et al.* 2010). With imbibition being the first step of seed germination, it has been hypothesized that a decrease in imbibition rate might reduce imbibitional injury associated with fast swelling seeds (VERTUCCI AND LEOPOLD 1987; WANG *et al.* 2010). Wang *et al.* 2011 measured imbibition by weighing water content of seeds at 24 and 48 h, and mapped QTL using multiple interval mapping (MIM) based on a linkage map of 135 SSR markers. The population studied consisted of 150 RILs which were cross from a salt-tolerant landrace (Jicaiqing) and a commercially valuable salt-susceptible *indica* accession, IR26.

Six QTLs were mapped on chromosomes 2, 3, 4, 8, 10 and 12 for IR under control conditions. The phenotypic variance explained by a single QTL ranged from 4.6 to 43.6% and the positive alleles could enhance IR by 21.6-86.5 mg/g. After comparing the QTLs identified at their two sampled time points (24 and 48 h), IR for 24 h was regulated by two QTLs (*qIR-4* and *qIR-12*), accounting for 40.9% of phenotypic variance. One major QTL *qIR-4* with a LOD score of 4.1 explained 22.9% of the phenotypic variance. On the other hand, another four QTLs (*qIR-2*, *qIR-3*, *qIR-8* and *qIR-10*) were responsible for the imbibition rate at the 48 h time point; these QTLs explained 64.5% of the total phenotypic variance, with one major QTL *qIR-10* (LOD = 2.6) having a R^2 = 43.6%. The expression of QTLs was developmentally regulated and four of

the six positive alleles of these QTLs originated from IR26. Comparing the two treatments (NaCl vs. control), only on QTL (*qIR-4*) was found for both conditions, which might play essential roles in seed imbibition. The frequency distribution of imbibition rate exhibited continuous and transgressive segregation, suggesting seed germination and IR are quantitative traits controlled by several genes (WANG *et al.* 2011).

An additional study in rice examined and mapped QTL for seedling vigor traits, comparing QTL co-localization with other traits in the literature (CHENG *et al.* 2013). They examined heterotrophic seedling growth (seedling dry weight; SDW), which is defined as a product of two key traits: the weight of the mobilized seed reserve (WMSR) and the conversion efficiency of the utilized seed reserve (seed reserve utilization efficiency; SRUE) into seedling tissue (16-17). While they identified upwards of 30 QTL, interestingly, several co-localized with imbibition rate QTL identified by Wang *et al.* 2011. For example, a QTL identified on chromosome 3 for SDW co-localized with *qIR-3*, an imbibition rate QTL, and *q7GR*, a germination rate QTL (ZHANG *et al.* 2005; WANG *et al.* 2011). The most robust imbibition rate QTL on chromosome 4, *qIR-4*, co-localized with both a germination percentage (GP) and SRUE QTL (ZHANG *et al.* 2005; WANG *et al.* 2011; CHENG *et al.* 2013). Lastly, on chromosome 12, an imbibition rate QTL co-localized with QTL for SDW, WMSR, and shoot weight (ZHANG *et al.* 2005; WANG *et al.* 2011; CHENG *et al.* 2013). These interesting results indicate, that contrary to evidence in *Medicago truncatula*, genomic regions which influence IR may also influence other important, later-stage germination QTL, indicating that imbibition rate may be a rapid, economically feasible phenotypic-proxy for breeders to select for to enhance germination and stand counts in other commercially valuable crops such as maize.

Imbibition, the first phase of germination, involves a rapid uptake of water by the dry seed, and has been shown to affect germination rate, overall stand count, and seedling vigor, all of which are important pre-requisites ultimately affecting optimum yield potential. The genetic architecture of maize seed traits is incredibly complex due to the lack of correspondence between transcriptional, proteomic, and metabolome levels attributable to dynamic post-transcriptional and post-translational regulation in maize seed tissues. Additionally, maize seed traits are influenced by the effects of seed production environment, seed parent genetic effects due to differential gene dosages in unique seed compartments, and/or parental imprinting of alleles, further adding to their genetic complexity. Hence, the goal of this thesis was to utilize of a systems-based 'omics' approach, integrating image-based phenomics, quantitative genetics, transcriptomics and proteomics to dissect the genetic architecture of seed imbibition and early germination in maize.

To accomplish these goals, we first created and evaluated the efficacy of a semi-automated, high-throughput image-based seed imbibition phenotyping platform across time to measure initial seed size, imbibition rate *K*, and percent increase in seed surface area 24 hours post-imbibing (%ISSA). Second, we identified sources of variance for these traits due to experimental design, seed production environmental effects, and genetic combining ability. Third, we evaluated the relationships among these seed imbibition traits, as well as their repeatabilities and heritabilities. Subsequently, the genetic architecture of quantitative variation in maize seed imbibition traits was dissected in a diverse maize population using genomic, transcriptomic, and proteomic datasets, identifying putative candidate genes and molecular pathways associated with maize seed imbibition. Finally, we also demonstrated a highly significant relationship of imbibition rate with radicle emergence time and total germination

percentage, indicating that our rapid phenotyping platform exhibits potential agronomic utility. In summary, this thesis constitutes a substantial effort towards establishing research and breeding strategies expected to improve early germination characteristics of maize.

LITERATURE CITED

Alexander, H., and R. Wulff, 1985 Experimental ecological genetics in Plantago: X. The effects of maternal temperature on seed and seedling characters in P. lanceolata. The Journal of Ecology**:** 271-282.

Antonovics, J., and J. Schmitt, 1986 Paternal and maternal effects on propagule size in Anthoxanthum odoratum. Oecologia 69**:** 277-282.

Attucci, S., J. P. Carde, P. Raymond, V. Saint-Gès, A. Spiteri *et al.*, 1991 Oxidative phosphorylation by mitochondria extracted from dry sunflower seeds. Plant physiology 95**:** 390-398.

Barrôco, R. M., K. Van Poucke, J. H. Bergervoet, L. De Veylder, S. P. Groot *et al.*, 2005 The role of the cell cycle machinery in resumption of postembryonic development. Plant Physiology 137**:** 127-140.

Barthole, G., L. Lepiniec, P. M. Rogowsky and S. Baud, 2012 Controlling lipid accumulation in cereal grains. Plant science 185**:** 33-39.

Baskin, C., and J. Baskin, 1998 Seeds. Ecology, biogeography, and evolution of dormancy and germination. Academic, New York.

BeMiller, J. N., 2007 *Carbohydrate chemistry for food scientists*. American Association of Cereal Chemists, Inc (AACC).

Bethke, P., S. J. Swanson, S. Hillmer and R. Jones, 1998 From storage compartment to lytic organelle: the metamorphosis of the aleurone protein storage vacuole. Annals of Botany 82**:** 399-412.

Bethke, P. C., I. G. Libourel, N. Aoyama, Y.-Y. Chung, D. W. Still *et al.*, 2007 The Arabidopsis aleurone layer responds to nitric oxide, gibberellin, and abscisic acid and is sufficient and necessary for seed dormancy. Plant Physiology 143**:** 1173-1188.

Bewley, J. D., 1997 Seed germination and dormancy. The Plant Cell 9**:** 1055.

Bewley, J. D., and M. Black, 1994 *Seeds*. Springer.

Biere, A., 1991a Parental effects in Lychnis flos-cuculi. I: Seed size, germination and seedling performance in a controlled environment. Journal of Evolutionary Biology 4**:** 447-465.

Biere, A., 1991b Parental effects in Lychnis flos-cuculi. II: Selection on time of emergence and seedling performance in the field. Journal of Evolutionary Biology 4**:** 467-486.

Blacklow, W., 1972 Mathematical description of the influence of temperature and seed quality on imbibition by seeds of corn (*Zea mays L.*). Crop Science 12**:** 643-646.

Boyer, C., and J. Shannon, 1986 Carbohydrate utilization in maize kernels.

Bradford, K. J., 1986 Manipulation of seed water relations via osmotic priming to improve germination under stress conditions. HortScience (USA).

Bradford, K. J., 1990 A water relations analysis of seed germination rates. Plant Physiology 94**:** 840-849.

Bradford, K. J., 1995 Water relations in seed germination. Seed development and germination 1**:** 351-396.

Bray, C. M., and C. E. West, 2005 DNA repair mechanisms in plants: crucial sensors and effectors for the maintenance of genome integrity. New Phytologist 168**:** 511-528.

Bruckler, L., 1983 Rôle des propriétés physiques du lit de semences sur l'imbibition et la germination. II. Contrôle expérimental d'un modèle d'imbibition des semences et possibilités d'applications. Agronomie 3**:** 223-232.

Cal, J., and R. Obendorf, 1972 Imbibitional chilling injury in *Zea mays L.* altered by initial kernel moisture and maternal parent. Crop Science 12**:** 369-373.

Calero, E., S. West and K. Hinson, 1981 Water Absorption of Soybean Seeds ans Associated Causal Factors. Crop Science 21**:** 926-933.

Carpita, N. C., M. W. Nabors, C. W. Ross and N. L. Petretic, 1979 The growth physics and water relations of red-light-induced germination in lettuce seeds. Planta 144**:** 225-233.

Chakraborty, R., and R. Kar, 2008 Differential water uptake kinetics in axes and cotyledons during seed germination of Vigna radiata under chilling temperature and cycloheximide treatment. Brazilian Journal of Plant Physiology 20**:** 277-284.

Chehab, E. W., C. Yao, Z. Henderson, S. Kim and J. Braam, 2012 Arabidopsis touch-induced morphogenesis is jasmonate mediated and protects against pests. Current Biology 22**:** 701-706.

Chen, F., P. Dahal and K. J. Bradford, 2001 Two tomato expansin genes show divergent expression and localization in embryos during seed development and germination. Plant Physiology 127**:** 928-936.

Chen, F., H. Nonogaki and K. J. Bradford, 2002 A gibberellin-regulated xyloglucan endotransglycosylase gene is expressed in the endosperm cap during tomato seed germination. Journal of Experimental Botany 53**:** 215-223.

Cheng, X., J. Cheng, X. Huang, Y. Lai, L. Wang *et al.*, 2013 Dynamic quantitative trait loci analysis of seed reserve utilization during three germination stages in rice.

Chung, D. S., and H. Pfost, 1967 Adsorption and desorption of water vapour by cereal grains and their products. Transactions of the ASAE 4**:** 549-551.

Cook, J. P., M. D. McMullen, J. B. Holland, F. Tian, P. Bradbury *et al.*, 2012 Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. Plant Physiology 158**:** 824-834.

Crowe, J., and L. Crowe, 1992 Membrane integrity in anhydrobiotic organisms: toward a mechanism for stabilizing dry cells, pp. 87-103 in *Water and Life*. Springer.

Dahal, P., and K. J. Bradford, 1994 Hydrothermal time analysis of tomato seed germination at suboptimal temperature and reduced water potential. Seed Science Research 4**:** 71-80.

De la Roche, I., D. Alexander and E. Weber, 1971 Inheritance of oleic and linoleic acids in *Zea mays L*. Crop Science 11**:** 856-859.

Debeaujon, I., and M. Koornneef, 2000 Gibberellin requirement for Arabidopsis seed germination is determined both by testa characteristics and embryonic abscisic acid. Plant physiology 122**:** 415-424.

Debeaujon, I., A. J. Peeters, K. M. Léon-Kloosterziel and M. Koornneef, 2001 The TRANSPARENT TESTA12 gene of Arabidopsis encodes a multidrug secondary transporter-like protein required for flavonoid sequestration in vacuoles of the seed coat endothelium. The Plant Cell 13**:** 853-871.

Dekkers, B. J., S. Pearce, R. van Bolderen-Veldkamp, A. Marshall, P. Widera *et al.*, 2013 Transcriptional dynamics of two seed compartments with opposing roles in Arabidopsis seed germination. Plant physiology 163**:** 205-215.

Demilly, D., M. Wagner, S. Brunel and C. Dürr, 2007 Computer vision for helping analysis of relationship between seed physical characteristics and germination and root elongation on Medicago truncatula, pp. in *28th ISTA Seed Symposium*.

Denny, F., 1917 Permeability of membranes as related to their composition. Botanical Gazette**:** 468-485.

Dias, P. M. B., S. Brunel-Muguet, C. Dürr, T. Huguet, D. Demilly *et al.*, 2011 QTL analysis of seed germination and pre-emergence growth at extreme temperatures in Medicago truncatula. Theoretical and Applied Genetics 122**:** 429-444.

Dickson, M., and M. Boettger, 1982 Heritability of semi-hard seed induced by low seed moisture in beans (Phaseolus vulgaris L.). J. Amer. Soc. Hort. Sci 107**:** 69.

Dubreucq, B., N. Berger, E. Vincent, M. Boisson, G. Pelletier *et al.*, 2000 The Arabidopsis AtEPR1 extensin-like gene is specifically expressed in endosperm during seed germination. The Plant Journal 23**:** 643-652.

Duke, S. H., G. Kakefuda and T. M. Harvey, 1983 Differential leakage of intracellular substances from imbibing soybean seeds. Plant Physiology 72**:** 919-924.

Eagles, H., and A. Hardacre, 1979 Genetic variation in maize (*Zea mays L*.) for germination and emergence at 10 C. Euphytica 28**:** 287-295.

Eckhoff, S. R., and M. R. Okos, 1989 Diffusion of gaseous sulfur-dioxide into corn grain. Cereal Chemistry 66**:** 30-33.

Ehrenshaft, M., and R. Brambl, 1990 Respiration and mitochondrial biogenesis in germinating embryos of maize. Plant physiology 93**:** 295-304.

Fan, L. T., P. S. Chu and J. A. Shellenberger, 1962 Volume increase of kernels of corn and sorghum accompanying absorption of liquid water. Biotechnology and Bioengineering 4**:** 311-322.

Fath, A., P. Bethke, J. Lonsdale, R. Meza-Romero and R. Jones, 2000 Programmed cell death in cereal aleurone, pp. 11-22 in *Programmed Cell Death in Higher Plants*. Springer.

Gardarin, A., C. Dürr and N. Colbach, 2011 Prediction of germination rates of weed species: relationships between germination speed parameters and species traits. Ecological Modelling 222**:** 626-636.

Gimeno-Gilles, C., E. Lelièvre, L. Viau, M. Malik-Ghulam, C. Ricoult *et al.*, 2009 ABA-mediated inhibition of germination is related to the inhibition of genes encoding cell-wall biosynthetic and architecture: modifying enzymes and structural proteins in Medicago truncatula embryo axis. Molecular plant 2**:** 108-119.

Graeber, K., A. Linkies, T. Steinbrecher, K. Mummenhoff, D. Tarkowská *et al.*, 2014 DELAY OF GERMINATION 1 mediates a conserved coat-dormancy mechanism for the temperature-and gibberellin-dependent control of seed germination. Proceedings of the National Academy of Sciences 111**:** E3571-E3580.

Groot, S., and C. Karssen, 1987 Gibberellins regulate seed germination in tomato by endosperm weakening: a study with gibberellin-deficient mutants. Planta 171**:** 525-531.

Groot, S. P., B. Kieliszewska-Rokicka, E. Vermeer and C. M. Karssen, 1988 Gibberellin-induced hydrolysis of endosperm cell walls in gibberellin-deficient tomato seeds prior to radicle protrusion. Planta 174**:** 500-504.

Grosh, G., and M. Milner, 1959 Water penetration and internal cracking in tempered wheat grains. Cereal Chem 36**:** 260-273.

Gubler, F., R. Kalla, J. K. Roberts and J. V. Jacobsen, 1995 Gibberellin-regulated expression of a myb gene in barley aleurone cells: evidence for Myb transactivation of a high-pI alpha-amylase gene promoter. The Plant Cell 7**:** 1879-1891.

Gummerson, R., 1986 The effect of constant temperatures and osmotic potentials on the germination of sugar beet. Journal of Experimental Botany**:** 729-741.

Gutterman, Y., 1992 Influences of daylength and red or far-red light during the storage of ripe Cucumis prophetarum fruits, on seed germination in light. Journal of arid environments 23**:** 443-449.

Gutterman, Y., 2012 *Seed germination in desert plants*. Springer Science & Business Media.

Hargin, K., W. Morrison and R. Fulcher, 1980 Triglyceride deposits in the starchy endosperm of wheat. Cereal chemistry.

Harper, J., and R. Benton, 1966 The behaviour of seeds in soil: II. The germination of seeds on the surface of a water supplying substrate. The Journal of Ecology**:** 151-166.

Harvey, B., and A. Oaks, 1974 The hydrolysis of endosperm protein in *Zea mays*. Plant physiology 53**:** 453-457.

Haughn, G., and A. Chaudhury, 2005 Genetic analysis of seed coat development in Arabidopsis. Trends in plant science 10**:** 472-477.

He, D., C. Han, J. Yao, S. Shen and P. Yang, 2011 Constructing the metabolic and regulatory pathways in germinating rice seeds through proteomic approach. Proteomics 11**:** 2693-2713.

Hedley, C. L., 2001 *Carbohydrates in grain legume seeds: Improving nutritional quality and agronomic characteristics*. CABI.

Heneen, W. K., G. Karlsson, K. Brismar, P.-O. Gummeson, S. Marttila *et al.*, 2008 Fusion of oil bodies in endosperm of oat grains. Planta 228**:** 589-599.

Herman, E. M., and B. A. Larkins, 1999 Protein storage bodies and vacuoles. The Plant Cell 11**:** 601-613.

Hoekstra, F., E. Golovina, A. Van Aelst and M. Hemminga, 1999 Imbibitional leakage from anhydrobiotes revisited. Plant, Cell & Environment 22**:** 1121-1131.

Howell, K. A., R. Narsai, A. Carroll, A. Ivanova, M. Lohse *et al.*, 2009 Mapping metabolic and transcript temporal switches during germination in rice highlights specific transcription factors and the role of RNA instability in the germination process. Plant physiology 149**:** 961-980.

Hueros, G., E. Gomez, N. Cheikh, J. Edwards, M. Weldon *et al.*, 1999 Identification of a promoter sequence from the BETL1gene cluster able to confer transfer-cell-specific expression in transgenic maize. Plant physiology 121**:** 1143-1152.

Hueros, G., S. Varotto, F. Salamini and R. D. Thompson, 1995 Molecular characterization of BET1, a gene expressed in the endosperm transfer cells of maize. The Plant Cell 7**:** 747-757.

Jacobsen, J., R. Knox and N. Pyliotis, 1971 The structure and composition of aleurone grains in the barley aleurone layer. Planta 101**:** 189-209.

Jerkovic, A., A. M. Kriegel, J. R. Bradner, B. J. Atwell, T. H. Roberts *et al.*, 2010 Strategic distribution of protective proteins within bran layers of wheat protects the nutrient-rich endosperm. Plant physiology 152**:** 1459-1470.

Jones, R. L., 1969 The fine structure of barley aleurone cells. Planta 85**:** 359-375.

Kantar, F., C. J. PILBEAM and P. D. Hebblethwaite, 1996 Effect of tannin content of faba bean (Vicia faba) seed on seed vigour, germination and field emergence. Annals of applied biology 128**:** 85-93.

Keeling, P. L., and A. M. Myers, 2010 Biochemistry and genetics of starch synthesis. Annual review of food science and technology 1**:** 271-303.

Kiesselbach, T. A., 1999 *The structure and reproduction of corn*. Cold spring harbor laboratory press.

Koornneef, M., L. Bentsink and H. Hilhorst, 2002 Seed dormancy and germination. Current opinion in plant biology 5**:** 33-36.

Koornneef, M., G. Reuling and C. Karssen, 1984 The isolation and characterization of abscisic acid-insensitive mutants of Arabidopsis thaliana. Physiologia Plantarum 61**:** 377-383.

Koornneef, M., and J. Van der Veen, 1980 Induction and analysis of gibberellin sensitive mutants in Arabidopsis thaliana (L.) Heynh. Theoretical and Applied genetics 58**:** 257-263.

Lacey, E., S. Smith and A. Case, 1997 Parental effects on seed mass: seed coat but not embryo/endosperm effects. American Journal of Botany 84**:** 1617-1617.

Lacey, E. P., 1996 Parental effects in *Plantago lanceolata* LI: a growth chamber experiment to examine pre-and postzygotic temperature effects. Evolution**:** 865-878.

Larkins, B. A., and W. J. Hurkman, 1978 Synthesis and deposition of zein in protein bodies of maize endosperm. Plant Physiology 62**:** 256-263.

Lee, K., R. Jones, A. Dalby and C. Tsai, 1976 Genetic regulation of storage protein content in maize endosperm. Biochemical genetics 14**:** 641-650.

Leeder, J., and I. Watt, 1974 The stoichiometry of water sorption by proteins. Journal of Colloid and Interface Science 48**:** 339-344.

Leleji, O., M. Dickson, L. Crowder and J. Bourke, 1972 Inheritance of crude protein percentage and its correlation with seed yield in beans, Phaseolus vulgaris L. Crop Science 12**:** 168-171.

Leopold, A. C., 1983 Volumetric Components of Seed Imbibition. Plant Physiology 73**:** 677-680.

Leopold, A. C., C. W. Vertucci, P. Stanwood and M. McDonald, 1989 Moisture as a regulator of physiological reaction in seeds. Seed moisture.**:** 51-67.

Leubner-Metzger, G., 2002 Seed after-ripening and over-expression of class I β-1, 3-glucanase confer maternal effects on tobacco testa rupture and dormancy release. Planta 215**:** 959-968.

Leubner-Metzger, G., C. Frundt, R. Vogeli-Lange and F. Meins Jr, 1995 Class I [beta]-1, 3-Glucanases in the Endosperm of Tobacco during Germination. Plant Physiology 109**:** 751-759.

Liptay, A., and P. Schopfer, 1983 Effect of water stress, seed coat restraint, and abscisic acid upon different germination capabilities of two tomato lines at low temperature. Plant Physiology 73**:** 935-938.

Lockhart, J. A., 1965 An analysis of irreversible plant cell elongation. Journal of theoretical biology 8**:** 264-275.

Marbach, I., and A. M. Mayer, 1974 Permeability of seed coats to water as related to drying conditions and metabolism of phenolics. Plant physiology 54**:** 817-820.

Marcus, A., J. Feeley and T. Volcani, 1966 Protein synthesis in imbibed seeds III. Kinetics of amino acid incorporation ribosome activation, and polysome formation. Plant physiology 41**:** 1167-1172.

Matakiadis, T., A. Alboresi, Y. Jikumaru, K. Tatematsu, O. Pichon *et al.*, 2009 The *Arabidopsis* abscisic acid catabolic gene CYP707A2 plays a key role in nitrate control of seed dormancy. Plant Physiology 149**:** 949-960.

Matilla, A., M. Gallardo and M. I. Puga-Hermida, 2005 Structural, physiological and molecular aspects of heterogeneity in seeds: a review. Seed Science Research 15**:** 63-76.

Mayer, A. M., and A. Poljakoff-Mayber, 1982 *The Germination of Seeds: Pergamon International Library of Science, Technology, Engineering and Social Studies*. Elsevier.

McDonald, M., C. Vertucci and E. Roos, 1988a Seed coat regulation of soybean seed imbibition. Crop Science 28**:** 987-992.

McDonald, M., C. Vertucci and E. Roos, 1988b Soybean seed imbibition: water absorption by seed parts. Crop Science 28**:** 993-997.

McDonald, M. B., 1994 Seed germination and seedling establishment. Physiology and determination of crop yield**:** 37-60.

Meyers, M. T., 1924 The influence of broken pericarp on the germination and yield of corn. Agronomy Journal 16**:** 540-550.

Mitchum, M. G., S. Yamaguchi, A. Hanada, A. Kuwahara, Y. Yoshioka *et al.*, 2006 Distinct and overlapping roles of two gibberellin 3-oxidases in *Arabidopsis* development. The Plant Journal 45**:** 804-818.

Moose, S. P., J. W. Dudley and T. R. Rocheford, 2004 Maize selection passes the century mark: a unique resource for 21st century genomics. Trends in plant science 9**:** 358-364.

Morohashi, Y., and J. D. Bewley, 1980 Development of mitochondrial activities in pea cotyledons during and following germination of the axis. Plant Physiology 66**:** 70-73.

Morohashi, Y., and M. Shimokoriyama, 1972 Physiological Studies on Germination of *Phaseolus mungo Seeds* I. Development of Respiration and Changes in the Contents of Constituents in the Early Stages of Germination. Journal of experimental botany 23**:** 45-53.

Müller, K., A. Linkies, R. A. Vreeburg, S. C. Fry, A. Krieger-Liszkay *et al.*, 2009 *In vivo* cell wall loosening by hydroxyl radicals during cress seed germination and elongation growth. Plant Physiology 150**:** 1855-1865.

Munir, J., L. A. Dorn, K. Donohue and J. Schmitt, 2001 The effect of maternal photoperiod on seasonal dormancy in *Arabidopsis thaliana* (Brassicaceae). American Journal of Botany 88**:** 1240-1249.

Murphy, J. B., and T. L. Noland, 1982 Temperature effects on seed imbibition and leakage mediated by viscosity and membranes. Plant physiology 69**:** 428-431.

Muthukumarappan, K., and S. Gunasekaran, 1994 Moisture Diffusivity of Corn Kernel Components During Adsorption Part II: Pericap. Transactions-American Society of Agricultural Engineers 37**:** 1269-1269.

Nakabayashi, K., M. Okamoto, T. Koshiba, Y. Kamiya and E. Nambara, 2005 Genome-wide profiling of stored mRNA in *Arabidopsis thaliana* seed germination: epigenetic and genetic regulation of transcription in seed. The plant journal 41**:** 697-709.

Nelson, D. C., J.-A. Riseborough, G. R. Flematti, J. Stevens, E. L. Ghisalberti *et al.*, 2009 Karrikins discovered in smoke trigger *Arabidopsis* seed germination by a mechanism requiring gibberellic acid synthesis and light. Plant Physiology 149**:** 863-873.

Nicolas, G., and J. Aldasoro, 1979 Activity of the pentose phosphate pathway and changes in nicotinamide nucleotide content during germination of seeds of *Cicer arietinum L*. Journal of Experimental Botany 30**:** 1163-1170.

Nieuwhof, M., F. Garretsen and J. Oeveren, 1989 Maternal and genetic effects on seed weight of tomato, and effects of seed weight on growth of genotypes of tomato (*Lycopersicon esculentum Mill.*). Plant breeding 102**:** 248-254.

Noggle, G. R., and G. J. Fritz, 1983 *Introductory plant physiology*. Prentice-Hall Inc.

Nomura, T., Y. Kono and T. Akazawa, 1969 Enzymic mechanism of starch breakdown in germinating rice seeds II. Scutellum as the site of sucrose synthesis. Plant Physiology 44**:** 765-769.

North, H., S. Baud, I. Debeaujon, C. Dubos, B. Dubreucq *et al.*, 2010 *Arabidopsis* seed secrets unravelled after a decade of genetic and omics-driven research. The Plant Journal 61**:** 971-981.

Oh, E., S. Yamaguchi, J. Hu, J. Yusuke, B. Jung *et al.*, 2007 PIL5, a phytochrome-interacting bHLH protein, regulates gibberellin responsiveness by binding directly to the GAI and RGA promoters in *Arabidopsis* seeds. The Plant Cell 19**:** 1192-1208.

Olsen, O.-A., 2001 Endosperm development: cellularization and cell fate specification. Annual review of plant biology 52**:** 233-267.

Parrish, J., and F. Bazzaz, 1985 Nutrient content of Abutilon theophrasti seeds and the competitive ability of the resulting plants. Oecologia 65**:** 247-251.

Passardi, F., C. Penel and C. Dunand, 2004 Performing the paradoxical: how plant peroxidases modify the cell wall. Trends in plant science 9**:** 534-540.

Platenkamp, G. A., and R. G. Shaw, 1993 Environmental and genetic maternal effects on seed characters in *Nemophila menziesii*. Evolution**:** 540-555.

Potts, H., J. Duangpatra, W. Hairston and J. Delouche, 1978 Some influences of hardseededness on soybean seed quality. Crop Science 18**:** 221-224.

Powell, A. A., 1986 Cell membranes and seed leachate conductivity in relation to the quality of seed for sowing. Journal of Seed Technology**:** 81-100.

Powell, A. A., and S. Matthews, 1978 The damaging effect of water on dry pea embryos during imbibition. Journal of Experimental Botany 29**:** 1215-1229.

Preston, J., K. Tatematsu, Y. Kanno, T. Hobo, M. Kimura *et al.*, 2009 Temporal expression patterns of hormone metabolism genes during imbibition of *Arabidopsis thaliana* seeds: a comparative study on dormant and non-dormant accessions. Plant and cell physiology 50**:** 1786-1800.

Purdy, J., and P. Crane, 1967 Influence of pericarp on differential drying rate in "mature" corn (*Zea mays L.*). Crop Science 7**:** 379-381.

Rajjou, L., K. Gallardo, I. Debeaujon, J. Vandekerckhove, C. Job *et al.*, 2004 The effect of α-amanitin on the *Arabidopsis* seed proteome highlights the distinct roles of stored and neosynthesized mRNAs during germination. Plant Physiology 134**:** 1598-1613.

Ramos, G., M. Pezet-Valdez, A. O'Connor-Sánchez, C. Placencia and R. C. Pless, 2004 Hydration rates for various types of Mexican maize based on single-kernel measurements. Cereal chemistry 81**:** 308-313.

Ramos, O. F., 2009 Physical properties, water absorption rate, equilibrium moisture content, and NIR composition of yellow, white, and specialty type maize hybrids, pp. PURDUE UNIVERSITY.

Ratkovic, S., and P. Pissis, 1997 Water binding to biopolymers in different cereals and legumes: proton NMR relaxation, dielectric and water imbibition studies. Journal of Materials Science 32**:** 3061-3068.

Ritchie, S. W., J. J. Hanway, G. O. Benson and J. Herman, 1993 How a corn plant develops, pp. Iowa State University of Science and Technology, Cooperative Extension Service.

Roach, D. A., and R. D. Wulff, 1987 Maternal effects in plants. Annual review of ecology and systematics**:** 209-235.

Robert, C., A. Noriega, Á. Tocino and E. Cervantes, 2008 Morphological analysis of seed shape in *Arabidopsis thaliana* reveals altered polarity in mutants of the ethylene signaling pathway. Journal of plant physiology 165**:** 911-919.

Ruan, R., J. Litchfield and R. Eckhoff, 1992 Simultaneous and Nondestructive Measurement of Transient Moisture Profiles and Structural Changes in Corn Kernels During Steeping Using Microscopic Nuclear Magnetic Resonance Imaging. Cereal Chem 69**:** 600-606.

Salon, C., P. Raymond and A. Pradet, 1988 Quantification of carbon fluxes through the tricarboxylic acid cycle in early germinating lettuce embryos. Journal of Biological Chemistry 263**:** 12278-12287.

Sánchez, R., and L. De Miguel, 1997 Phytochrome promotion of mannan-degrading enzyme activities in the micropylar endosperm of *Datura ferox* seeds requires the presence of the embryo and gibberellin synthesis. Seed Science Research 7**:** 27-34.

Sano, N., H. Permana, R. Kumada, Y. Shinozaki, T. Tanabata *et al.*, 2012 Proteomic analysis of embryonic proteins synthesized from long-lived mRNAs during germination of rice seeds. Plant and Cell Physiology 53**:** 687-698.

Schoettle, A. W., and A. C. Leopold, 1984 Solute leakage from artificially aged soybean seeds after imbibition. Crop Science 24**:** 835-838.

Schopfer, P., 2006 Biomechanics of plant growth. American journal of botany 93**:** 1415-1425.

Schopfer, P., and C. Plachy, 1985 Control of Seed Germination by Abscisic Acid III. Effect on Embryo Growth Potential (Minimum Turgor Pressure) and Growth Coefficient (Cell Wall Extensibility) in *Brassica napus L.* Plant Physiology 77**:** 676-686.

Schwaegerle, K. E., and D. A. Levin, 1990 Quantitative genetics of seed size variation in *Phlox.* Evolutionary Ecology 4**:** 143-148.

Seo, M., A. Hanada, A. Kuwahara, A. Endo, M. Okamoto *et al.*, 2006 Regulation of hormone metabolism in *Arabidopsis* seeds: phytochrome regulation of abscisic acid metabolism and abscisic acid regulation of gibberellin metabolism. The Plant Journal 48**:** 354-366.

Seo, M., E. Nambara, G. Choi and S. Yamaguchi, 2009 Interaction of light and hormone signals in germinating seeds. Plant molecular biology 69**:** 463-472.

Shao, S., C. J. Meyer, F. Ma, C. A. Peterson and M. A. Bernards, 2007 The outermost cuticle of soybean seeds: chemical composition and function during imbibition. Journal of experimental botany 58**:** 1071-1082.

Shaykewich, C., 1973 Proposed method for measuring swelling pressure of seeds prior to germination. Journal of Experimental Botany 24**:** 1056-1061.

Shewry, P. R., and N. G. Halford, 2002 Cereal seed storage proteins: structures, properties and role in grain utilization. Journal of Experimental Botany 53**:** 947-958.

Shull, C. A., 1920 Temperature and rate of moisture intake in seeds. Botanical Gazette**:** 361-390.

Shull, C. A., and S. Shull, 1924 Temperature coefficient of absorption in seeds of corn. Botanical Gazette**:** 262-279.

Singh, L., and H. H. Hadley, 1972 Maternal and cytoplasmic effects on seed protein content in soybeans, *Glycine max (L.)* Merrill. Crop Science 12**:** 583-585.

Sliwinska, E., G. W. Bassel and J. D. Bewley, 2009 Germination of *Arabidopsis thaliana* seeds is not completed as a result of elongation of the radicle but of the adjacent transition zone and lower hypocotyl. Journal of experimental botany 60**:** 3587-3594.

Sreenivasulu, N., B. Usadel, A. Winter, V. Radchuk, U. Scholz *et al.*, 2008 Barley grain maturation and germination: metabolic pathway and regulatory network commonalities and differences highlighted by new MapMan/PageMan profiling tools. Plant physiology 146**:** 1738-1758.

Stewart, A., H. Nield and J. N. Lott, 1988 An investigation of the mineral content of barley grains and seedlings. Plant physiology 86**:** 93-97.

Stiles, I. E., 1948 Relation of water to the germination of corn and cotton seeds. Plant physiology 23**:** 201.

Stroshine, R., J. Tuite and P. Crane, 1987 Effect of kernel physical properties on shelled-corn thin-layer drying rates. American Society of Agricultural Engineers (Microfiche collection)(USA).

Swift, J., and T. O'Brien, 1970 Vascularization of the scutellum of wheat. Australian Journal of Botany 18**:** 45-53.

Swift, J., and T. O'brien, 1972 The fine structure of the wheat scutellum before germination. Australian Journal of Biological Sciences 25**:** 9-22.

Syarief, A., R. Gustafson and R. V. Morey, 1984 Moisture diffusion coefficients for yellow-dent corn components. Paper-American Society of Agricultural Engineers (USA).

Tatum, L., and M. Zuber, 1943 Germination of maize under adverse conditions. J. Amer. Soc. Agron 35**:** 48-59.

Taylor, A., J. Prusinski, H. Hill and M. Dickson, 1992 Influence of seed hydration on seedling performance. HortTechnology 2**:** 336-344.

Thompson, R. D., G. Hueros, H.-A. Becker and M. Maitz, 2001 Development and functions of seed transfer cells. Plant Science 160**:** 775-783.

Toh, S., A. Imamura, A. Watanabe, K. Nakabayashi, M. Okamoto *et al.*, 2008 High temperature-induced abscisic acid biosynthesis and its role in the inhibition of gibberellin action in *Arabidopsis* seeds. Plant physiology 146**:** 1368-1385.

True, R. H., 1893 On the development of the caryopsis. Botanical Gazette**:** 212-226.

Tully, R. E., M. E. Musgrave and A. C. Leopold, 1981 The seed coat as a control of imbibitional chilling injury. Crop Science 21**:** 312-317.

Varner, J., 1964 Gibberellic acid controlled synthesis of α-amylase in barley endosperm. Plant physiology 39**:** 413.

Vertucci, C., and A. Leopold, 1986 Physiological activities associated with hydration level in seeds. Membranes, Metabolism and Dry Organisms. Comstock Publishing Associates, Ithaca, NY**:** 35-49.

Vertucci, C. W., 1989 The effects of low water contents on physiological activities of seeds. Physiologia Plantarum 77**:** 172-176.

Vertucci, C. W., and A. C. Leopold, 1983 Dynamics of imbibition by soybean embryos. Plant physiology 72**:** 190-193.

Vertucci, C. W., and A. C. Leopold, 1987 Water binding in legume seeds. Plant physiology 85**:** 224-231.

Wang, H., C. Offler and J. Patrick, 1994 Nucellar projection transfer cells in the developing wheat grain. Protoplasma 182**:** 39-52.

Wang, Z.-f., J.-f. Wang, Y.-m. Bao, Y.-y. Wu, S. Xuan *et al.*, 2010 Inheritance of rice seed germination ability under salt stress. Rice Science 17**:** 105-110.

Wang, Z., J. Wang, Y. Bao, Y. Wu and H. Zhang, 2011 Quantitative trait loci controlling rice seed germination under salt stress. Euphytica 178**:** 297-307.

Waters, A. J., P. Bilinski, S. R. Eichten, M. W. Vaughn, J. Ross-Ibarra *et al.*, 2013 Comprehensive analysis of imprinted genes in maize reveals allelic variation for imprinting and limited conservation with other species. Proceedings of the National Academy of Sciences 110**:** 19639-19644.

Waters, A. J., I. Makarevitch, S. R. Eichten, R. A. Swanson-Wagner, C.-T. Yeh *et al.*, 2011 Parent-of-origin effects on gene expression and DNA methylation in the maize endosperm. The Plant Cell 23**:** 4221-4233.

Watson, S., and E. Sanders, 1961 Steeping studies with corn endosperm sections. Cereal Chem 38**:** 22-33.

Watson, S., P. White and L. Johnson, 2003 Description, development, structure, and composition of the corn kernel. Corn: chemistry and technology**:** 69-106.

Weitbrecht, K., K. Müller and G. Leubner-Metzger, 2011 First off the mark: early seed germination. Journal of experimental botany 62**:** 3289-3309.

Williams, W., and R. McGibbon, 1980 Genetic control of seed weight and seed oil in *Lupinus albus* and *Lupinus mutabilis*. Zeitschrift fur Pflanzenzuchtung 84**:** 329-334.

Wolf, M., C. Buzan, M. M. MacMASTERS and C. Rist, 1952 Structure of the mature corn kernel. 1. Gross anatomy and structural relationships. Cereal Chemistry 29**:** 321-333.

Wolf, M., U. Khoo and H. Seckinger, 1969 Distribution and subcellular structure of endosperm protein in varieties of ordinary and high-lysine maize. Cereal Chemistry 46**:** 253-263.

Woodstock, L., 1988 Seed imbibition: a critical period for successful germination. Journal of Seed Technology**:** 1-15.

Wyatt, J., 1977 Seed coat and water absorption properties of seed of near-isogenic snap bean lines differing in seed coat color. Journal of the American Society for Horticultural Science 102**:** 478-480.

Yaklich, R., E. Vigil and W. Wergin, 1986 Pore development and seed coat permeability in soybean. Crop science 26**:** 616-624.

Yamauchi, Y., M. Ogawa, A. Kuwahara, A. Hanada, Y. Kamiya *et al.*, 2004 Activation of gibberellin biosynthesis and response pathways by low temperature during imbibition of *Arabidopsis thaliana* seeds. The Plant Cell 16**:** 367-378.

Young, T. E., and D. R. Gallie, 1999 Analysis of programmed cell death in wheat endosperm reveals differences in endosperm development between cereals. Plant molecular biology 39**:** 915-926.

Zamski, E., and A. A. Schaffer, 1996 *Photoassimilate distribution in Plants and crops: source-sink relationships*. Marcel Dekker Inc.

Zeng, C.-L., J.-b. Wang, A.-H. Liu and X.-M. Wu, 2004 Seed coat microsculpturing changes during seed development in diploid and amphidiploid *Brassica* species. Annals of Botany 93**:** 555-566.

Zhang, Z.-H., S.-B. Yu, T. Yu, Z. Huang and Y.-G. Zhu, 2005 Mapping quantitative trait loci (QTLs) for seedling-vigor using recombinant inbred lines of rice (*Oryza sativa L.*). Field Crops Research 91**:** 161-170.

Zheng, Y., and Z. Wang, 2010 Current opinions on endosperm transfer cells in maize. Plant cell reports 29**:** 935-942.

Zhou, X., S. Yan, C. Sun, S. Li, J. Li *et al.*, 2015 A Maize Jasmonate Zim-Domain Protein, ZmJAZ14, Associates with the JA, ABA, and GA Signaling Pathways in Transgenic Arabidopsis. PloS one 10**:** e0121824.

Acknowledgement of co-authorship for Chapter 2

Authors: S. Stelpflug[1], N. Miller[2], E. Spalding[2], and S. Kaeppler[2]

Affiliations: [1]Department of Agronomy, University of Wisconsin, 1575 Linden Drive, Madison, WI 53706; [2]Department of Botany, University of Wisconsin, 132 Birge Hall, 430 Lincoln Drive, Madison, WI 53706

Author's contributions: SS, NM, ES, and SK conceived and designed the study. SS collected the data, performed statistical analyses and wrote the manuscript. NM designed the phenotypic assay and the programming code to analyze and export phenotypic data. All authors read and approved the final manuscript.

CHAPTER 2. EVALUATION OF SEED LOT, MATERNAL EFFECTS, AND COMBINING ABILITY EFFECTS ON MAIZE SEED IMBIBIBITION TRAITS

ABSTRACT

Imbibition, the first phase of germination, involves a rapid uptake of water by the dry seed, and has been shown to affect germination rate, overall stand count, and seedling vigor, all of which are important pre-requisites ultimately affecting optimum yield potential. Understanding genetic and non-genetic components of variation among maize inbreds for imbibition rate and other seed quality will be useful towards developing breeding strategies. Here, we report two independent studies designed to assess the effects of environment, maternal parent, and genetic combining ability on maize seed imbibition rate (*K*), percent increase in seed surface area (%ISSA), and seed size via analysis of variance. The first experiment assayed seed of 25 diverse maize inbreds produced in three distinct years to evaluate the effect of environment. The second experiment was a complete 9x9 diallel experiment with reciprocal crosses to estimate general combining ability (GCA), specific combining ability (SCA), and reciprocal genetic effects (RE). Although the effect of seed source on seed traits was significant, high repeatabilities and heritabilities indicated that the observed phenotypic variation in both experiments was largely of genetic origin, reaffirming the potential for genetic improvement and association mapping of early germination traits in maize. Analysis of variance of GCA, SCA, and RE were significant for all traits measured in the full-diallel, excluding the SCA effects of %ISSA. The quadratic components of the fixed-effects for SCA and RE were higher than those for GCA for traits *K* and %ISSA, which indicates higher significance of non-additive genetic effects in respect to *K* and %ISSA. Hence, breeders should consider the SCA and reciprocal effects of seed parents when

attempting to optimize imbibition rates in maize to improve seedling vigor and uniformity of stands.

2.1 INTRODUCTION

The success of seed germination and the establishment of a normal seedling are fundamental in the propagation of plant species grown from seed, having both economic and ecological importance. Rapid germination rate and overall emergence is an important pre-requisite for obtaining satisfactory stands, seedling establishment, and seedling vigor, all of which ultimately affect optimum yield potential in maize (SHAW AND LOOMIS 1950; MARTIN *et al.* 1988).

Seed germination begins when dry seeds come in contact with water under favorable conditions. Phase I of this triphasic process, imbibition, involves a rapid uptake of water by the dry seed. The imbibition period offers opportunity as well as hazard; seeds may be primed for increased vigor by imbibing and then drying back, and high membrane permeability during early imbibition may facilitate insertion of germination-promoting and anti-pathogen chemicals into seed tissues (WOODSTOCK 1988). However, rapid uptake of water may cause imbibitional injury. Alternatively, slow imbibition rates may extend the duration of adverse conditions incurred by the seed: cold temperatures may cause chilling injury, anaerobic conditions may lead to an accumulation of toxins, leaching of cellular constituents into the soil may promote microbial attack, or the seed might be damaged by soil drying due to depletion of water from the soil surface (WOODSTOCK 1988). Imbibition rate has been shown to positively or negatively affect germination rate and stand counts in snap peas, depending on soil water conditions (TAYLOR *et al.* 1992). Thus, it is essential to investigate the nature of the genetic effects that affect imbibition

rate and other seed quality traits with respect to germination and seedling vigor to fine-tune maize performance.

Variation in imbibition rate in maize could be due to a combination of the genetic constitution of the seed, genetic constitution of the maternal parent, and non-genetic environmental effects. This is due to the fact that seeds consist of tissue representing two generations: the caryopsis, which is a fusion of the fruit pericarp and the seed coat, is diploid maternal tissue, while the triploid endosperm and diploid embryo are composed of tissue representing the next generation. Second, seed traits are influenced by at least two environments: while maturing, they are attached to the mother plant and are influenced by the maternal/ seed production environment; subsequently, they are dispersed to the germination environment.

Maternal effects, the phenotypic contribution of the mother to her offspring beyond the equal chromosomal contribution expected from each parent, are anticipated to be significant in maize seeds. Genetic maternal effects are potentially considerable, because (1) the pericarp is genotypically maternal, (2) the endosperm and aleurone layer have a larger maternal (2n) than paternal nuclear contribution (n), (3) the endosperm and embryo cytoplasms in maize (including the mitochondrial and chloroplast DNAs) are maternally inherited, and (4) certain genes are imprinted and silenced, depending on maternal or paternal transmission (WATERS *et al.* 2011; WATERS *et al.* 2013). The endosperm and aleurone contain enzymes important for germination and is also the source of nutrients for the developing embryo (HARVEY AND OAKS 1974). As a consequence of the differential dosage of male and female genes, the female parent may have a more important role in determining the characteristics of this nutrient source. Therefore, it is not surprising that relatively large maternal effects have been found in seed phenotypes of higher

plants (ALEXANDER AND WULFF 1985; ANTONOVICS AND SCHMITT 1986; NIEUWHOF *et al.* 1989; SCHWAEGERLE AND LEVIN 1990; BIERE 1991a; BIERE 1991b).

Genetic studies have shown a large maternal effect on seed size and germination percentage in maize (EAGLES AND HARDACRE 1979). For instance, fatty acid content of seeds was shown to be significantly influenced by the maternal parent in maize hybrids (DE LA ROCHE *et al.* 1971). Additionally, maternal effects on seed mineral composition have been documented for several additional species, such as for oil content in *Lupinus* (WILLIAMS AND MCGIBBON 1980) and protein content in dry beans and soybeans (LELEJI *et al.* 1972; SINGH AND HADLEY 1972). In the case of dry beans and soybeans, these differences did not persist into the F_2 generation, suggesting that these are non-cytoplasmic maternal effects. In addition to influencing the rate or timing of germination, maternal genetic effects can also influence the sensitivity of seeds to environmental conditions; for instance, in maize, a maternal effect was found for sensitivity chilling injury during the imbibition phase of germination (CAL AND OBENDORF 1972).

The final class of maternal effects is the interaction of the maternal plant with the seed production environment. Seed production effects are expected to be large in certain seed traits because of the seeds' attachment to the mother plant during their development and growth. These influences may occur via structure or physiology. The tissues immediately surrounding the developing embryo and endosperm are all maternal. These tissues, the integuments of the ovule and the wall of the ovary, eventually form accessory seed structures which are important determinants of seed dispersal, dormancy, and germination traits, and variation in these traits can carry over to influence the mature phenotype of an individual. The aforementioned seed traits vary with environmental conditions such as temperature (LACEY 1996), photoperiod

(GUTTERMAN 1992; MUNIR *et al.* 2001), nutrient availability (PARRISH AND BAZZAZ 1985), and soil moisture conditions (GUTTERMAN 2012). Environmental control of germination acts through the seed coat, the endosperm, nutrient and hormone supply (BIERE 1991a; PLATENKAMP AND SHAW 1993; LACEY *et al.* 1997; BASKIN AND BASKIN 1998). While a considerable number of quantitative genetic studies related to seed characteristics have been conducted (ROACH AND WULFF 1987), very few have jointly analyzed the effects of maternal environment, maternal genetic effects, and their interactions (PLATENKAMP AND SHAW 1993).

Despite the importance of developing maize varieties with high germination and emergence rates, no studies have reported the predominant genetic effects controlling seed imbibition rate in maize. As maize varieties are primarily produced and sold as hybrids, studying their combining ability via full-diallel analysis (GRIFFING 1956) may yield valuable information to facilitate the selection of the best seed parents to create enhanced hybrids with optimized imbibition rates based on their general combining ability (GCA), specific combining ability (SCA), and reciprocal/maternal effects (RE). Here, we report two independent studies designed to assess the effects of the seed production environment and the genetic combining ability effects on maize seed imbibition via analysis of variance: (1) a seed source experiment of 25 diverse maize inbreds whose maternal parents were grown from three distinct years to evaluate the effect of the seed production environment, and (2) a 9x9 full-diallel experiment with reciprocal crosses to estimate GCA, SCA, and reciprocal genetic effects. We also examined phenotypic correlations between seed imbibition traits and reported their observed repeatability and heritability. This study is part of a larger project in which we investigated the quantitative genetic aspects of maize kernel imbibition and its contribution to germination rate, as described in Chapter 3.

2.2 MATERIALS AND METHODS

Germplasm

To evaluate general combining ability (GCA), specific combining ability (SCA) and reciprocal effects, we created a 9x9 full-diallel using maize inbreds A663, B14A, B73, DKHBA1, Ky226, H121, Mo5, Mo17, NK740 as parents to produce 72 hybrid offspring (including reciprocals) at West Madison Agricultural Research Station during the summer of 2014 (parental inbreds listed in **Appendix A1**). For the evaluation of putative seed source effects, we evaluated 25 inbreds (**Appendix A1**) from seed lots produced via inbred increase during the summers of 2010, 2012, and 2013.

Phenotypic Assay of Kernel Imbibition

Uniform, flat maize kernels were aligned with a square grid template, with one inch spacing between kernels, and were placed on square bioassay dishes (Corning #43111, 245mm x 245mm) filled with 150 mL of 1% agar (Difco brand) per petri dish. A 1% agar solution was prepared either the day of the assay or the day before in batches of 1L, using 10 grams of agar per 1L of deionized H_20 and allowed to cool and solidify. Nine kernels per genotype (occupying one column each) per row were placed on the agar petri dishes, with ten genotypes placed per agar petri dish (experimental set-up pictured in **Figure 1**). Upon completion of kernel plating, kernels were covered with one layer of green paper (8 x 11 in, Avery Paper Company) to serve as background contrast for image analysis, and were subsequently covered with two layers of paper towel, evenly soaked with 20 mLs of deionized H_20 to saturate the paper towels to imbibe the kernels. Upon wetting, agar plates were immediately transferred to flat-bed scanners (Epson

Perfection V700 photo flatbed scanner; 12x20x6 in) for image acquisition and analysis (one petri dish per scanner, six scanners running per day). Kernels were imbibed at room temperature (20ºC). Images of swelling kernels were acquired every ten minutes for 24 hours at a resolution of 1200 dpi using a custom Unix script, and images were exported to MATLAB for analysis (Miller et al., 2015 unpublished) to derive three different kernel phenotypes: initial seed size (measured in pixels), percent increase in seed surface area after 24 hours (%ISSA), and imbibition rate, given as a constant termed "*K*." %ISSA was calculated using the formula %ISSA = $(Area_{24Hr}\text{-}Area_{initial})/Area_{initial.}$ Imbibition rate constant *K* was calculated by fitting the raw kernel area at every time point to a previously defined equation: $Area_{max}(1 - e^{-Kt}) = Area(t)$, solving for *K* (LEOPOLD 1983). An example schematic of these three imbibition phenotypes for two contrasting genotypes is depicted in **Figure 2**.

Experimental Design and Statistical Analysis

Analysis of variance (ANOVA) for the seed source experiment was performed using the *anova* function of the R software (R DEVELOPMENT CORE TEAM 2013). Experimental design of the seed source experiment included three replications total, blocked in runs of time, with subsamples of nine seeds representing a genotype within a replicate. After filtering of outlier values with studentized residuals > |3|, mean values of the nine kernels were calculated and used for analysis of each replication. The linear model specified with the *lm* function in R included the fixed effects of genotype, seed source, column, and replication. The following model was used: Y = genotype + source + column + replication + error. Boxplots and histograms were also drawn using the R software.

Analysis of variance of general combining ability (GCA), specific combining ability (SCA), and reciprocal combining ability (RCA) or reciprocal effects of the full-diallel were calculated using the *diallel1* function based on Griffing's diallel model 1, where both the genotypes and replications are treated as fixed effects (GRIFFING 1956) contained within the *plantbreeding* package of R software (R DEVELOPMENT CORE TEAM 2013). Parents and replicates are considered fixed in this analysis as they are not random representatives of a defined population but specifically chosen inbred lines. This method is appropriate when the varieties sampled include the parent varieties and the variety crosses (including reciprocals), as was the case in our study. Griffing's analysis using a fixed model, which unlike the random model does not require an assumption of zero epistatic effects or an independent distribution of genes among parents, as only the GCA and SCA genetic effects are estimated (GRIFFING 1956). Sprague and Tatum (1942) first defined GCA as "the average performance of a line in hybrid combinations" and SCA as "those cases in which certain combinations do relatively better or worse than would be expected on the basis of the average performance of the lines crossed."

The 81 genotypes (72 reciprocal hybrids, nine inbred parents) were phenotypically evaluated as described previously, using four replicates blocked in time with subsamples of nine seeds representing a genotype within a replicate. After filtering of outlier values with studentized residuals > |3|, mean values of the nine kernels were calculated and used for analysis of each replication. Data was analyzed using a genetic model which includes both additive and dominance effects (GRIFFING 1956) per the following model:

$$y_{ijl} = \mu + g_i + g_j + s_{ij} + r_{ij} + b_l + e_{ijl},$$

where y_{ij} = the mean trait value of the hybrid F_1s and reciprocals (i,j = 1, 2, …p); μ is the general population mean; g_i, g_j = the effect of general combining ability (GCA) effect for the *i*th and *j*th parents, respectively; s_{ij} = the effects of specific combining ability (SCA) for crosses between the *i*th and *j*th parents such that $s_{ij} = s_{ji}$; r_{ij} = the reciprocal effect (RE) that measures the differences provided by the parent of *i* or *j* when used as a male or female in cross *ij*, such that $r_{ij} = -r_{ji}$, b_l is the effect of replicate l = 1..4; e_{ijkl} = the average experimental error/ residual effects associated with the *ijl*th individual observation. Significance of GCA, SCA, and RCA effect ≠ zero was tested via *F*-test as $F = GCA^2/var(\widehat{g_i})$, $SCA^2/var(\widehat{s_{ii}})$ or $SCA^2/var(\widehat{s_{ij}})$, and $RCA^2/var(\widehat{r_{ij}})$, with degrees of freedom equal to the number of observations (df =240). Heat maps depicting SCA and RCA effects were constructed using the *pheatmap* function using R software (R DEVELOPMENT CORE TEAM 2013). The quadratic forms (analogous to a variance component, but referring to a fixed effect) associated with the GCA, SCA, and RCA effects were estimated by the method of moments based on expected mean squares of fixed effects according to the following formulas (GRIFFING 1956; HUNG AND HOLLAND 2012):

$$\widehat{\varphi_{GCA}} = \frac{MS_{GCA} - MS_E}{2p},\ \widehat{\varphi_{SCA}} = MS_{SCA} - MS_E,\ \widehat{\varphi_{RCA}} = \frac{MS_{RCA} - MS_E}{2},$$

wherein p = number of parents, MS_{GCA} = mean square of GCA, MS_{SCA} = mean square of SCA, MS_{RCA}= mean square of RCA, MS_E= mean square of error. The relative importance of GCA and SCA was assessed as the ratio: $\frac{2\varphi_{GCA}}{2\varphi_{GCA} + \varphi_{SCA}}$ (BAKER 1978). As this ratio approaches unity (equal to one), GCA becomes more important, and the predictability of the performance of a specific hybrid could be based on the GCA effects of the parents in isolation (BAKER 1978).

Analyses of variance (ANOVA) for the repeatability and heritability calculations were performed using the *lmer* function of the R package *lme4 (*BATES *et al.* 2015*)*. The mixed effect model used in the ANOVA and heritability calculations derived from the seed source experiment included the overall mean and seed source year as fixed effects, with genotype, genotype-by-source interaction, and replication as random effects. The *lmer* function of the R package *lme4 (*BATES *et al.* 2015*)* was utilized to estimate variance components with the restricted maximum likelihood method (REML). Model assumptions (normality of residuals, homogeneous variance of residuals and normality of random effects) were assessed in each model. Repeatabilities for all traits evaluated in the full-diallel and seed source experiments were calculated with a mixed model, including genotype effect and replication effect (i.e. block effect), both as random effects. Repeatability was calculated for the full-diallel experiment as follows: $= \frac{\sigma_G^2}{\sigma_G^2 + \sigma_e^2/r}$, where σ_G^2 is the genotypic variance, σ_e^2 is the error variance, and r is the number of technical replications. Pair-wise Pearson's product moment correlations between phenotypic traits were performed using the *cor.test* function of R (R DEVELOPMENT CORE TEAM 2013).

2.3 RESULTS AND DISCUSSION

Analysis of variance to deduce seed source effects

In order to isolate the relative contribution of environmental seed production effects from genetic influences on seed imbibition traits (*K*, %ISSA, seed size), we selected 25 representative maize inbred lines (listed in **Appendix A1**) to evaluate the effect of seed lots produced during different environments from the summers of 2010, 2012, and 2013 via analysis of variance (ANOVA). The *F*-test from the ANOVA revealed significant effects ($p < 0.001$) for all model inputs: genotype, seed source, genotype-by-source interaction, replication, and column placement on the scanner. Genotype had the largest F-value for traits *K* and seed size, whereas the largest *F*-value for %ISSA was due to the effect of replication. While the effect of seed source on imbibition rate *K* was significant ($p < 0.001$), the mean square of seed source was ~3.6 times less than that of genotype (**Table 1**). Overlaying the phenotypic distributions of *K* by seed source indicates that the newest seed source (2013) has proportionally more individuals with higher *K* values, whereas seed produced during 2012, a year which experienced extreme drought in Wisconsin, had lower *K* values (**Figure 3A**). However, the box-plot distributions and mean of *K* by seed source indicate that the differences in magnitude are minimal, with no systematic trends (**Figure 3B**). Pearson's correlations of mean genotypic *K* values across seed sources were moderately high and statistically significant ($r = 0.62$ for 2010 vs. 2012, $r = 0.67$ for 2010 vs. 2013, and $r = 0.63$ for 2012 vs. 2013, $p < 0.001$ for all comparisons). This effect was non-systematic, with different genotypes being superior in different years of seed production. We conclude that the effect of seed source on *K* is significant but non-systematic due to the observation that *K* values from all seed sources are significantly positively correlated, and that the magnitude of variance explained by genotype is much larger than the variance explained by

source (**Table 1**). Hence, significant genetic variance exists for this trait, making *K* amenable to genetic mapping studies. Importantly, previous seed trait QTL mapping experiments of identical RIL populations across multiple seed sources in *Arabidopsis* have found unique seed source-dependent QTLs (MOORE *et al.* 2013), similar to QTLs which are dependent genotype-by-environment interactions (JANSEN *et al.* 1995). While it was not feasible for our genetic studies given a lack of inventory of genotypes for certain seed sources, it would be informative to map seed QTLs for RIL populations grown out over multiple seed sources in future maize mapping experiments to identify more robust seed trait QTLs.

Analysis of variance to determine experimental and image-acquisition biases

While our phenotypic measurements were performed in controlled, uniform environments absent of environmental effects typically observed for field-based phenotyping, it remained important for us to evaluate systematic sources of bias in our phenotypic data. For instance, the effect of both column and replication was significant for all traits measured ($p < 0.001$), particularly for seed size (**Table 1 and Appendix A2**). The effect of replication caused a decrease in *K* values with increasing replication, due to a bias in selecting more uniform seeds first for phenotypic analysis (**Appendix A3**). The observed column effect was due to lower *K* values of kernels placed in the outermost columns of the scanner (columns one and nine) (**Appendix A2 and A3**). We attributed this to a "fish eye" effect, where kernels furthest from the focal point of the scanner on the edge appeared larger- this trend is more apparent when looking at the effect of column on seed size (**Appendix A2**). Therefore, the observed column and replication effects were taken into account for all subsequent phenotypic and genetic analysis related to maize kernel imbibition traits.

Phenotypic diversity and the relationships amongst imbibition traits

We compared and contrasted the phenotypic distributions of *K*, %ISSA, and seed size measured from two separate experiments (**Figure 4**): 1) the seed source experiment described above and 2) a 9x9 full-diallel experiment which is described further in the subsequent section of this manuscript. The mean and standard deviation of *K* was 0.0178 ± 0.005 for the seed source experiment and 0.021 ± 0.03 for the full-diallel experiment (**Figure 4A**), indicating that the full-diallel population exhibited both higher imbibition rates and more variation in general. *K* values from both the seed source and diallel experiments are higher than the previously reported swelling coefficient *K* (0.157) of maize, which was based on changes in seed volume (LEOPOLD 1983). However, Leopold (1983) did not sample maize genotypes for this calculation, so our calculated values may be more representative of typical maize swelling coefficients. Seed size mean values were 137,821 pixels ± 25,458 and 154,261 pixels ± 23,894 for the seed source and diallel experiments, respectively, while %ISSA mean values ranged from 20.2% ± 0.025 (seed source) to 21.1% ± 0.006 (**Figure 4B and 4C**). Interestingly, %ISSA and seed size phenotypic distributions exhibited larger standard deviations in the seed source experiment, whereas *K* exhibited a larger standard deviation value in the full-diallel experiment, perhaps due to the enhanced values of certain hybrid crosses in the population.

The relationships between *K*, %ISSA, and seed size were tested by Pearson's correlation analysis for both experiments (**Table 2**). The only trait correlation that was significant in both experiments was between *K* and %ISSA (r = -0.259 and -0.434, $p < 0.001$) for the seed source and diallel experiments, respectively, indicating that seeds which reach a larger total area after 24H of imbibition tend to have a slower initial imbibition rate *K*. Additionally, imbibition rate *K* exhibited a minor, but significantly positive correlation to seed size in the full-diallel experiment

(r = 0.104, p < 0.001) but was not significantly correlated in the seed source experiment (r = -0.015). This correlation is smaller in magnitude than a previously reported correlation between seed face area and water absorption in maize (r = 0.31, p < 0.05) (RAMOS 2009), but both studies in maize are in contrast to results in a *Medicago truncatula* study which found a significantly negative correlation between seed size and imbibition rate (r = -0.65, p < 0.001) (DIAS *et al.* 2011). These results indicate that imbibition rate is not substantially influenced by seed size in our assay, which is not very surprising given that initial seed size is accounted for in the calculation of K (LEOPOLD 1983). Additionally, the lack of substantial correlations amongst traits indicates that they are, for the most part, under separate genetic control. Considering that the diallel experiment sampled more kernels (n=2916) than were sampled for the seed source experiment (n=1896), it would be expected that correlations from the diallel experiment are more robust.

Repeatability and heritability for kernel imbibition traits

Repeatability (R) is broadly defined as the correlation between repeated measurements over time in the same individual and is a frequently used measurement in plant breeding because it sets the upper limit of heritability (FALCONER AND MACKAY 1996). Repeatability was high for K and seed size measured within the diallel population ($R \geq 80\%$) (**Table 3**). The large values for repeatability are indicative of the underlying genetic control of imbibition rate and seed size. On the other hand, %ISSA showed a considerably lower repeatability (R = 62.7%) indicating that this trait is comparatively more influenced by experimental error and/or the environment. As expected, repeatability values for K, %ISSA, and seed size from the seed source experiment (R = 60.0, 45.3 %, and 75.6%, respectively) were lower than repeatability values for each trait from the diallel experiment, which may be relatively upwardly biased due to the fact that they were

estimated from a single seed source in a uniform environment (**Table 3**). Nevertheless, genotypic mean squares for seed size and *K* were two and 3.5-fold higher than the mean squares attributed to seed source, respectively (**Table 4**), indicating that genetic variance controls the majority of observed phenotypic variance observed in our population relative to the variance explained by seed source and error. This observation supports that that these seed imbibition traits are suitable for genetic mapping and breeding studies.

Full-diallel combining ability analysis

The analysis of variance for the average performance of 72 hybrids along with their parents in 9 x 9 full-diallel crosses for all imbibition traits studied is presented in **Table 4**. The F-test revealed significant differences between genotypes supporting substantial natural genetic variation within the population studied ($p < 0.001$). The effects of GCA, SCA, and RCA were significant for all evaluated traits, excluding SCA effects of %ISSA, indicating that additive, non-additive and maternal effects are all involved in the genetic control of imbibition rate *K* and seed size (**Table 4**). Using diallel analysis in tropical maize lines to study seed quality, previous researchers have shown that GCA, SCA, and RE all significantly affect germination time to emergence and stand counts in sand bed vigor tests (MOTERLE *et al.* 2011), consistent with the hypothesis that imbibition, a component phase of germination, may be under similar genetic control as other previously reported germination traits.

The estimates of quadratic components associated with the fixed-effect mean squares of GCA, SCA, and RE indicated that non-additive effects had a greater impact on all maize seed and imbibition traits than additive effects (**Table 6**), with the magnitude of effects being RE > SCA > GCA for *K* and %ISSA. These results are consistent with multiple diallel experiments in

maize, which found that non-additive effects were more important for the genetic control of seed quality based on germination timing and cold vigor tests based on estimates of quadratic components associated with combining ability fixed effects (GOMES *et al.* 2000; MOTERLE *et al.* 2012; CABRAL *et al.* 2013). It is worth noting than an opposite trend has been observed in other maize diallel studies where the researchers observed that additive genetic effects had a larger influence on emergence rate and germination percentage (ANTUNA *et al.* 2003; CERVANTES ORTIZ *et al.* 2006), although these authors used only the mean squares associated with GCA and SCA effects of each trait and did not use the corresponding estimated quadratic components, which is the basis for the differing conclusions. Our results would have mirrored these latter results as well had we used an equivalent evaluation of raw mean squares, with the magnitude of effects being GCA>RCA>SCA for all traits.

General combining ability effects

The estimations of parental phenotypic means and GCA main effects, defined as the average performance of a parent among all crosses, which estimates a line's breeding value, were determined for each parent (SPRAGUE AND TATUM 1942) (**Figure 5**). Overall, statistical tests of GCA effects ≠ zero via *F*-test were significant ($p < 0.05$) for seven, five, and nine parents from the diallel for traits *K*, %ISSA, and seed size, respectively. Based on GCA effects, the best inbred parents for producing hybrid seed with faster imbibition rate *K* were DKHBA1, Mo17, and Ky226 ($p < 0.001$), whereas H121, Mo5, and B73 had the lowest GCA values (indicating slowest imbibition rates for *K* (**Figure 5**). Alternatively, the best inbreds for producing hybrid seed with the highest swelling capacity (%ISSA) were genotypes H121, Mo5 and Ky226, with

parents A663 and Mo17 causing the largest decrease in %ISSA. GCA effects related to seed size were highly significant ($p < 0.001$) for all nine parents, with NK740, Mo17, and B14A contributing the highest GCA values which enhanced seed size in their progenies; genotypes Mo5, Ky226, and A663 had the largest negative GCA effects related to seed size in their hybrid progenies (**Figure 5**). While the GCA quadratic components were the least significant of the three genetic effects dictating imbibition traits in this population (**Table 5**), they were nonetheless highly significant. All of the parents (excluding parent A663) of the five hybrids with the highest imbibition rates also had favorable positive GCA values towards *K*. This suggests that, although imbibition tends to be at least partly due to dominance and the effect of the mother plant, hybrid production will require crossing inbred lines that both have high imbibition values, assuming this is agronomically favorable. Additionally, the majority of the parents (seven of nine) had highly significant GCA effects in respect to *K*, supporting a strong additive genetic component in this set of lines.

Specific combining ability effects

Specific combining ability (SCA) effects, which indicate the differences in the effect of a specific cross from the average general combining ability of the parents, have been shown to correspond to genetic effects due to dominance, which are often indicative of heterosis, the superior performance of hybrid progenies relative to both parental lines (HAYMAN 1958; GARDNER AND EBERHART 1966; FALCONER AND MACKAY 1996). SCA effects of specific crosses relative to *K* are depicted in **Figure 6**. It is desirable to know the magnitude of effects of specific combinations in hybrid-based breeding schemes to optimize cultivar performance, which is the case for maize. Of all 36 hybrid combinations, 11 crosses (31%) had SCA effects which significantly differed from zero, with five and six crosses yielding significantly positive and

negative SCA effects, respectively. The crosses A663 x Ky226, Mo17 x Ky226, and DKHBA1 x NK740 showed highly significant ($P < 0.001$) and positive SCA effects for imbibition rate *K*, indicating that these crosses would be the best combinations to produce hybrids with enhanced and superior *K* values relative to the GCA effects of their parents in isolation (**Figure 7**). Alternatively, the crosses Mo17 x H121, NK740 x A663, and B73 x Ky226 show significant and negative SCA effects for imbibition rate, indicating that these are specific crosses would be expected to have slow imbibition rates (**Figure 6**). Interestingly, the cross B73 x Mo17 also exhibits significant and positive SCA effects (**Figure 6**). This is worth noting as B73 and Mo17 are lines from the stiff-stalk and non-stiff stalk heterotic pools, respectively, which have been selected to maximize heterotic performance (STUBER *et al.* 1992), indicating that in addition to yield, imbibition and early germination traits may also be highly heterotic when combining genotypes from opposing heterotic pools. Additionally, four of the nine self-pollinations of parental inbreds yielded significant SCA effects. Three self-pollinated parents (H121 ($p < 0.001$), B14A ($p < 0.01$), and NK740 ($p < 0.01$)) had positive SCA effects toward *K*, indicating that these lines imbibe faster when self-pollinated relative to any of their hybrid progenies, whereas self-pollinated parent Ky226 exhibited negative SCA, indicating that any of the hybrid progeny of this genotype imbibe faster than their Ky226 parent does. This observation is not unsurprising, as SCA effects may deviate from the "expected' value, which is the sum of the general combining abilities of its two parental lines (FALCONER AND MACKAY 1996). Non-coincidentally, heterosis has also been shown to enhance radicle emergence in maize hybrids (GUO *et al.* 2013). Proteomic analysis conducted on imbibing embryos between a hybrid and its parental lines indicated that 47% (63/134) protein spots displayed non-additive expression patterns in imbibed seed embryos, with 54.6% of non-additively accumulated proteins displaying

above or equal to the level of the higher parent expression patterns (GUO *et al.* 2013). These results indicate that non-additive genetic effects should be considered when creating hybrid maize varieties with enhanced seed quality and germination characteristics.

Reciprocal effects

The reciprocal effect (RE) quadratic components for all the traits measured were higher than those for SCA indicating that there is a significant difference between using a line as a male or female in the same cross (**Table 6**). These results corroborate previous observations in both tropical and popcorn maize, which found greater effects for RE than SCA for germination percentage and emergence rate in sand bed tests (MOTERLE *et al.* 2011; CABRAL *et al.* 2013). From a biological perspective, it could be that reciprocal effects dictate the largest proportion of genetic effects related to kernel imbibition traits due to the fact that the genotype (aa) of the pericarp is completely under maternal control. Indeed, the permeability and morphology of the seed coat and/or pericarp has been shown to be a major factor controlling the rate of water uptake by seeds (MAYER AND POLJAKOFF-MAYBER 1982; VALENTI *et al.* 1989). Removal of the pericarp from the maize kernel has been shown to cause a two-order magnitude increase in the moisture diffusion coefficient of seeds ($\sim 10^{-9}$ vs. $\sim 10^{-7}$) (MUTHUKUMARAPPAN AND GUNASEKARAN 1994). Intriguingly, a study of maize hybrids revealed that water absorption rate was highly correlated to the top ($r = 0.60$, $p < 0.0001$) and middle abgerminal side ($r = 0.41$, $p < 0.0001$) thickness of the pericarp (RAMOS 2009), which supports the hypothesis that significant maternal effects on *K* may be a caused by differences in pericarp structure.

RE of specific crosses relative to *K* are depicted in **Figure 7**. Of all 36 hybrid combinations, the majority of crosses (20, or 55.6%) exhibited reciprocal combining ability

(RCA) effects which significantly differed from zero when used as a female, with sixteen and four crosses yielding significantly positive and negative RCA effects, respectively. The crosses A663♀ x Ky226♂, Mo17♀ x H121♂, and NK740♀ x Mo5♂ showed the most significant ($p < 0.001$) highest positive RCA effects for imbibition rate *K* (**Figure 7**). Alternatively, the crosses Mo5♀ x DKHBA1♂ ($P < 0.001$), Mo5♀ x B73♂ ($p < 0.001$), and H121♀ x DKHBA1♂ ($p < 0.01$) exhibited the most negative RCA effects for *K* (**Figure 7**). Furthermore, two non-stiff stalk parents, NK740 and Mo17, which are 96% genetically related according to RNA-seq SNPs derived from seedling tissue (HIRSCH *et al.* 2014), contribute significant and positive RCA effects when used as females for eight different hybrid combinations, with crosses to Mo5, H121, and DKHBA1 being significantly positive for both parental crosses. Hybrid combinations that exhibit significantly positive RCA effects which are unique to each parent include Mo17♀ x B73♂ and NK740♀ x Ky226♂ (**Figure 7**). Interestingly, his indicates that Mo17 and NK740 are both ideal parents to use as females as opposed to males during hybrid seed production in order to optimize imbibition rates of their progenies, which is counter-intuitive given that these genotypes belong to the non-stiff stalk maize heterotic group typically used as males during seed production. Hence, as reciprocal effects imparted by mother plants are significant and far-reaching in respect to *K*, these effects should also be considered when selecting seed parents for hybrid production to fine-tune imbibition and germination characteristics in maize breeding.

2.4 CONCLUSIONS

We have generated a novel high-throughput phenotyping platform for maize which simultaneously assays imbibition rate *K*, overall seed water uptake (%ISSA), and seed size more accurately, faster, and with greater resolution than previously available methods. Further experiments will be necessary to determine if this phenotyping platform will have any utility towards predicting agronomically valuable germination traits in maize. Using two experiments to assay both the effects of seed production environment and general combining ability on imbibition traits in maize, we have partitioned and identified sources and magnitudes of variation related to GCA, SCA, reciprocal effects, seed production environment, seed sampling, and experimental design. These results will help to inform future experiments studying early germination and imbibition traits both in maize and other species. Both the seed source and full-diallel experiments indicated that substantial phenotypic variation exists for imbibition rate *K* and seed size. Analysis of variance of the seed source experiment revealed that the seed production environment had significant effects on all traits measured. However, the effects were non-systematic, with genotypic variances being substantially larger than genotype-by-source interaction variances for *K* and seed size. The relatively medium to high heritability and repeatability estimates reported for the seed and imbibition traits evaluated in the study additionally indicate that the phenotypic variation observed was largely of genetic origin. Likewise, the combining ability analysis conducted in our study revealed that the mean squares for GCA effects made significant and important contributions to hybrid variation for all traits studied, excluding the SCA effects relative to %ISSA. Therefore, all evidence points to high genetic control of these traits, indicating that there is significant potential for genetic mapping and improvement via selection and breeding of early germination traits in maize. Correlation

analysis between traits revealed modest to weak relationships amongst K, %ISSA, and seed size, indicating that these traits do not highly influence each other and are most likely under mostly unique genetic control.

The quadratic components of the fixed-effects for SCA and reciprocal effects were higher than those for GCA for traits *K* and %ISSA, which indicates higher significance of non-additive genetic effects in respect to imbibition traits *K* and %ISSA. The significant effect of seed parent on imbibition rate could be due to the fact that the rate of imbibition is highly influenced by the permeability and thickness of the pericarp (MAYER AND POLJAKOFF-MAYBER 1982; VALENTI *et al.* 1989; MUTHUKUMARAPPAN AND GUNASEKARAN 1994; RAMOS 2009). Hence, in addition to GCA, breeders should consider the SCA and reciprocal effects of seed parents when attempting to enhance early germination traits in maize. Overall, the results of these experiments have provided novel insights into the sources of environmental, maternal, and genotypic effects governing maize imbibition, which will help breeders optimize the performance of early germination traits in maize, map and identify genes underlying these traits, and aid in spurring additional hypotheses towards unraveling the biological nature of imbibition rate in maize

LITERATURE CITED

Alexander, H., and R. Wulff, 1985 Experimental ecological genetics in Plantago: X. The effects of maternal temperature on seed and seedling characters in P. lanceolata. The Journal of Ecology**:** 271-282.

Antonovics, J., and J. Schmitt, 1986 Paternal and maternal effects on propagule size in Anthoxanthum odoratum. Oecologia 69**:** 277-282.

Antuna, O., F. Rincón, E. Gutiérrez, N. A. Ruiz and L. Bustamante, 2003 Componentes genéticos de caracteres agronómicos y de calidad fisiológica de semillas en líneas de maíz.

Baker, R., 1978 Issues in diallel analysis. Crop Science 18**:** 533-536.

Baskin, C., and J. Baskin, 1998 Seeds. Ecology, biogeography, and evolution of dormancy and germination. Academic, New York.

Bates, D., B. Bolker and S. Walker, 2015 lme4: Linear mixed-effects models using S4 classes., pp. R Core Team.

Biere, A., 1991a Parental effects in Lychnis flos-cuculi. I: Seed size, germination and seedling performance in a controlled environment. Journal of Evolutionary Biology 4**:** 447-465.

Biere, A., 1991b Parental effects in Lychnis flos-cuculi. II: Selection on time of emergence and seedling performance in the field. Journal of Evolutionary Biology 4**:** 467-486.

Cabral, P. D. S., A. T. d. Amaral Júnior, H. D. Vieira, J. S. Santos, I. L. d. J. Freitas *et al.*, 2013 Genetic effects on seed quality in diallel crosses of popcorn. Ciência e Agrotecnologia 37**:** 502-511.

Cal, J., and R. Obendorf, 1972 Imbibitional chilling injury in *Zea mays L.* altered by initial kernel moisture and maternal parent. Crop Science 12**:** 369-373.

Cervantes Ortiz, F., G. García de los Santos, A. Carballo-Carballo, D. Bergvinson, J. L. Crossa *et al.*, 2006 Análisis dialélico para caracteres de vigor de semilla y de plántula en genotipos de maíz tropical. Agricultura técnica en México 32**:** 77-87.

De la Roche, I., D. Alexander and E. Weber, 1971 Inheritance of oleic and linoleic acids in *Zea mays L*. Crop Science 11**:** 856-859.

Dias, P. M. B., S. Brunel-Muguet, C. Dürr, T. Huguet, D. Demilly *et al.*, 2011 QTL analysis of seed germination and pre-emergence growth at extreme temperatures in Medicago truncatula. Theoretical and Applied Genetics 122**:** 429-444.

Eagles, H., and A. Hardacre, 1979 Genetic variation in maize (*Zea mays L.*) for germination and emergence at 10 C. Euphytica 28**:** 287-295.

Falconer, D. S., and T. F. C. Mackay, 1996 *Introduction to quantitative genetics*. Prentice Hall, New York, New York, USA, Pearson Education Limited.

Gardner, C., and S. Eberhart, 1966 Analysis and interpretation of the variety cross diallel and related populations. Biometrics**:** 439-452.

Gomes, M., E. Von Pinho, R. G. Von Pinho and M. Vieira, 2000 Estimativas da capacidade de combinação de linhagens de milho tropical para qualidade fisiológica de sementes. Ciência e Agrotecnologia 24**:** 41-49.

Griffing, B., 1956 Concept of general and specific combining ability in relation to diallel crossing systems. Australian Journal of Biological Sciences 9**:** 463-493.

Guo, B., Y. Chen, G. Zhang, J. Xing, Z. Hu *et al.*, 2013 Comparative proteomic analysis of embryos between a maize hybrid and its parental lines during early stages of seed germination. PLoS One.

Gutterman, Y., 1992 Influences of daylength and red or far-red light during the storage of ripe Cucumis prophetarum fruits, on seed germination in light. Journal of arid environments 23**:** 443-449.

Gutterman, Y., 2012 *Seed germination in desert plants*. Springer Science & Business Media.

Harvey, B., and A. Oaks, 1974 The hydrolysis of endosperm protein in Zea mays. Plant physiology 53**:** 453-457.

Hayman, B., 1958 The theory and analysis of diallel crosses. II. Genetics 43**:** 63.

Hirsch, C. N., J. M. Foerster, J. M. Johnson, R. S. Sekhon, G. Muttoni *et al.*, 2014 Insights into the maize pan-genome and pan-transcriptome. The Plant Cell Online 26**:** 121-135.

Hung, H.-Y., and J. B. Holland, 2012 Diallel analysis of resistance to Fusarium ear rot and fumonisin contamination in maize. Crop Science 52**:** 2173-2181.

Jansen, R., J. Van Ooijen, P. Stam, C. Lister and C. Dean, 1995 Genotype-by-environment interaction in genetic mapping of multiple quantitative trait loci. Theoretical and Applied Genetics 91**:** 33-37.

Lacey, E., S. Smith and A. Case, 1997 Parental effects on seed mass: seed coat but not embryo/endosperm effects. American Journal of Botany 84**:** 1617-1617.

Lacey, E. P., 1996 Parental effects in Plantago lanceolata LI: a growth chamber experiment to examine pre-and postzygotic temperature effects. Evolution**:** 865-878.

Leleji, O., M. Dickson, L. Crowder and J. Bourke, 1972 Inheritance of crude protein percentage and its correlation with seed yield in beans, Phaseolus vulgaris L. Crop Science 12**:** 168-171.

Leopold, A. C., 1983 Volumetric Components of Seed Imbibition. Plant Physiology 73**:** 677-680.

Martin, B., O. Smith and M. O'neil, 1988 Relationships between laboratory germination tests and field emergence of maize inbreds. Crop Science 28**:** 801-805.

Mayer, A. M., and A. Poljakoff-Mayber, 1982 *The Germination of Seeds: Pergamon International Library of Science, Technology, Engineering and Social Studies*. Elsevier.

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

Moterle, L., A. Braccini, C. Scapim, R. Pinto, L. Goncalves *et al.*, 2011 Combining ability of tropical maize lines for seed quality and agronomic traits. Genet Mol Res 10**:** 2268-2278.

Moterle, L. M., A. de Lucca, C. A. Scapim, R. J. B. Pinto, L. S. A. Gonçalves *et al.*, 2012 Combining ability of popcorn lines for seed quality and agronomic traits. Euphytica 185**:** 337-347.

Munir, J., L. A. Dorn, K. Donohue and J. Schmitt, 2001 The effect of maternal photoperiod on seasonal dormancy in Arabidopsis thaliana (Brassicaceae). American Journal of Botany 88**:** 1240-1249.

Muthukumarappan, K., and S. Gunasekaran, 1994 Moisture Diffusivity of Corn Kernel Components During Adsorption Part II: Pericap. TRANSACTIONS-AMERICAN SOCIETY OF AGRICULTURAL ENGINEERS 37**:** 1269-1269.

Nieuwhof, M., F. Garretsen and J. Oeveren, 1989 Maternal and genetic effects on seed weight of tomato, and effects of seed weight on growth of genotypes of tomato (Lycopersicon esculentum Mill.). Plant breeding 102**:** 248-254.

Parrish, J., and F. Bazzaz, 1985 Nutrient content of Abutilon theophrasti seeds and the competitive ability of the resulting plants. Oecologia 65**:** 247-251.

Platenkamp, G. A., and R. G. Shaw, 1993 Environmental and genetic maternal effects on seed characters in Nemophila menziesii. Evolution**:** 540-555.

R Development Core Team, 2013 R: A language and environment for statistical computing., pp. R Foundation for Statistical Computing, Vienna, Austria.

Ramos, O. F., 2009 Physical properties, water absorption rate, equilibrium moisture content, and NIR composition of yellow, white, and specialty type maize hybrids, pp. PURDUE UNIVERSITY.

Roach, D. A., and R. D. Wulff, 1987 Maternal effects in plants. Annual review of ecology and systematics**:** 209-235.

Schwaegerle, K. E., and D. A. Levin, 1990 Quantitative genetics of seed size variation inPhlox. Evolutionary Ecology 4**:** 143-148.

Shaw, R. H., and W. Loomis, 1950 Bases for the prediction of corn yields. Plant Physiology 25**:** 225.

Singh, L., and H. H. Hadley, 1972 Maternal and cytoplasmic effects on seed protein content in soybeans, Glycine max (L.) Merrill. Crop Science 12**:** 583-585.

Sprague, G. F., and L. A. Tatum, 1942 General versus specific combining ability in single crosses of corn. Journal of the American Society for Agronomy**:** 923-932.

Stuber, C. W., S. E. Lincoln, D. Wolff, T. Helentjaris and E. Lander, 1992 Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. Genetics 132**:** 823-839.

Taylor, A., J. Prusinski, H. Hill and M. Dickson, 1992 Influence of seed hydration on seedling performance. HortTechnology 2**:** 336-344.

Valenti, G. S., L. Melone, M. Ferro and A. Bozzini, 1989 Comparative studies on testa structure of hard-seeded and soft-seeded varieties of Lupinus angustifolius L.(Leguminosae) and on mechanisms of water entry. Seed science and technology 17**:** 563-581.

Waters, A. J., P. Bilinski, S. R. Eichten, M. W. Vaughn, J. Ross-Ibarra *et al.*, 2013 Comprehensive analysis of imprinted genes in maize reveals allelic variation for imprinting and limited conservation with other species. Proceedings of the National Academy of Sciences 110**:** 19639-19644.

Waters, A. J., I. Makarevitch, S. R. Eichten, R. A. Swanson-Wagner, C.-T. Yeh *et al.*, 2011 Parent-of-origin effects on gene expression and DNA methylation in the maize endosperm. The Plant Cell 23**:** 4221-4233.

Williams, W., and R. McGibbon, 1980 Genetic control of seed weight and seed oil in Lupinus albus and Lupinus mutabilis. Zeitschrift fur Pflanzenzuchtung 84**:** 329-334.

Woodstock, L., 1988 Seed imbibition: a critical period for successful germination. Journal of Seed Technology**:** 1-15.

Table 1. Analysis of variance (fixed-effects model) of seed and early germination traits from multiple seed sources. Genotypes represent a random sample of inbreds from the WiDiv diversity panel (*n*=25) (listed in **Appendix A1**). Seed sources represent inbred seed increases from summer nurseries spanning 2010, 2012, and 2013. The analysis of variance was performed on the mean phenotypic values of nine kernels per replication, omitting outliers with studentized residuals </> | 3 |. Residuals represent the interaction of all terms included in the model.

		Mean Squares			***F-value***		
Sources of variance	**df**	***K***	**%ISSA**	**Seed Size**	***K***	**%ISSA**	**Seed Size**
Genotype	24	0.0157	0.0093	3.59E+10	43.8 ***	15.0 ***	205.7 ***
Seed Source	2	0.00036	0.0038	1.82E+10	12.2 ***	6.1 **	104.2 ***
Column	8	0.00078	0.0053	2.09E+09	6.6 ***	8.6 ***	12.0 ***
Replication	2	0.00028	0.016	4.96E+09	9.6 ***	25.2 ***	28.4 ***
Residuals	1897	0.0282	0.00062	1.75E+08	---	---	---

*** $p < 0.001$, ** $p < 0.01$, *K*: imbibition rate constant, %ISSA: % increase of seed surface area after 24 hours, df: degrees of freedom.

Table 2. Pair-wise Pearson's correlations between imbibition rate, percentage increase in seed surface area after 24 H imbibition, and seed size of maize kernels for both the seed source and full-diallel experiments. ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, *ns*: not significant ($p > 0.05$); *r*: Pearson's correlation coefficient, *n*: number of observations in the experiment, *K*: imbibition rate constant, %ISSA: percentage increase of seed surface area after 24 hours imbibition.

	r	
Pair-wise Phenotypic Correlation	***Seed Source*** (*n*=1853)	***Diallel*** (*n*=2585)
K vs. %ISSA	-0.259 ***	-0.434 ***
K vs. Seed Size	-0.015 *ns*	0.104 ***
%ISSA vs. Seed Size	0.116 ***	0.013 *ns*

Table 3. Variance and repeatability estimates for imbibition rate, percentage increase in seed surface area after 24 H imbibition, and seed size of maize kernels for both the seed source and full-diallel experiments. Genotypes used for both experiments are listed in **Appendix A1**.Genotypic variance (σ^2_G) was significant for all phenotypic traits ($p < 0.001$). *R*: repeatability, σ^2_e: residual variance, *K*: imbibition rate constant, %ISSA: percentage increase of seed surface area after 24 hours imbibition.

Experiment	***Seed Source***			***Full-diallel***		
Trait	σ^2_G	σ^2_e	***R* (%)**	σ^2_G	σ^2_e	***R* (%)**
K	0.00286	0.0038	60.0	2.07E-05	1.83E-05	81.9
%ISSA	0.0103	0.025	45.3	0.0004184	0.0009923	62.7
Seed Size	21250	13753	75.6	559901671	128765665	94.5

Table 4. Full-diallel analysis of variance for imbibition rate, percentage increase in seed surface area after 24 H imbibition, and seed size of maize kernels. Diallel analysis was performed according to Griffing's model I (fixed-effects) (GRIFFING 1956). Genotypes used for the analysis include nine inbred parents (A663, B14A, B73, DKHBA1, H121, Ky226, Mo5, Mo17, and NK740) and all pairwise hybrid progeny, including reciprocal crosses (*n*=72). The analysis of variance was performed on the mean phenotypic values of nine kernels per replication, omitting outliers with studentized residuals </> |3|. Residuals represent the interaction of all terms included in the model.

		Mean Squares			***F-value***		
Sources of variance	**df**	***K***	**%ISSA**	**Seed Size**	***K***	**%ISSA**	**Seed Size**
Genotypes	80	0.00009072	0.001653	2227657938	13.8 ***	3.2 ***	99.3 ***
Replications	3	0.00039494	0.012393	271874493	60.1 ***	23.7 ***	12.1 ***
GCA	8	0.000090411	0.001165	3420528753	55.1 ***	8.9 ***	610.1 ***
SCA	36	0.000009727	0.000223	75994338	5.9 ***	1.7 *ns*	13.5 ***
RCA	36	0.000020584	0.000437	401475905	12.5 ***	3.4***	71.6 ***
Error	240	0.000001642	0.000131	5606875	---	---	---

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, *ns* = not significant. *K*: imbibition rate constant, %ISSA: percentage increase of seed surface area after 24 hours of imbibition, GCA: general combining ability, SCA: specific combining ability, RCA: reciprocal combining ability, df: degrees of freedom.

Table 5. Quadratic components associated with the fixed-effects of general combining ability (GCA), specific combining ability (SCA), and reciprocal effects (RE) for imbibition rate (*K*), percent increase in seed surface area (%ISSA), and seed size from the full-diallel experiment. The relative importance of general and specific combining ability in determining progeny performance was assessed by the ratio 2φg / 2φg + φs (BAKER 1978).

Estimates of quadratic components	*K*	%ISSA	Seed Size
GCA (φ_g)	4.9E-06	5.7E-05	1.9E+08
SCA (φ_s)	8.1E-06	9.2E-05	7.0E+07
RE (φ_{re})	9.5E-06	1.5E-04	2.0E+08
2φg / 2φg + φs	0.55	0.56	0.84

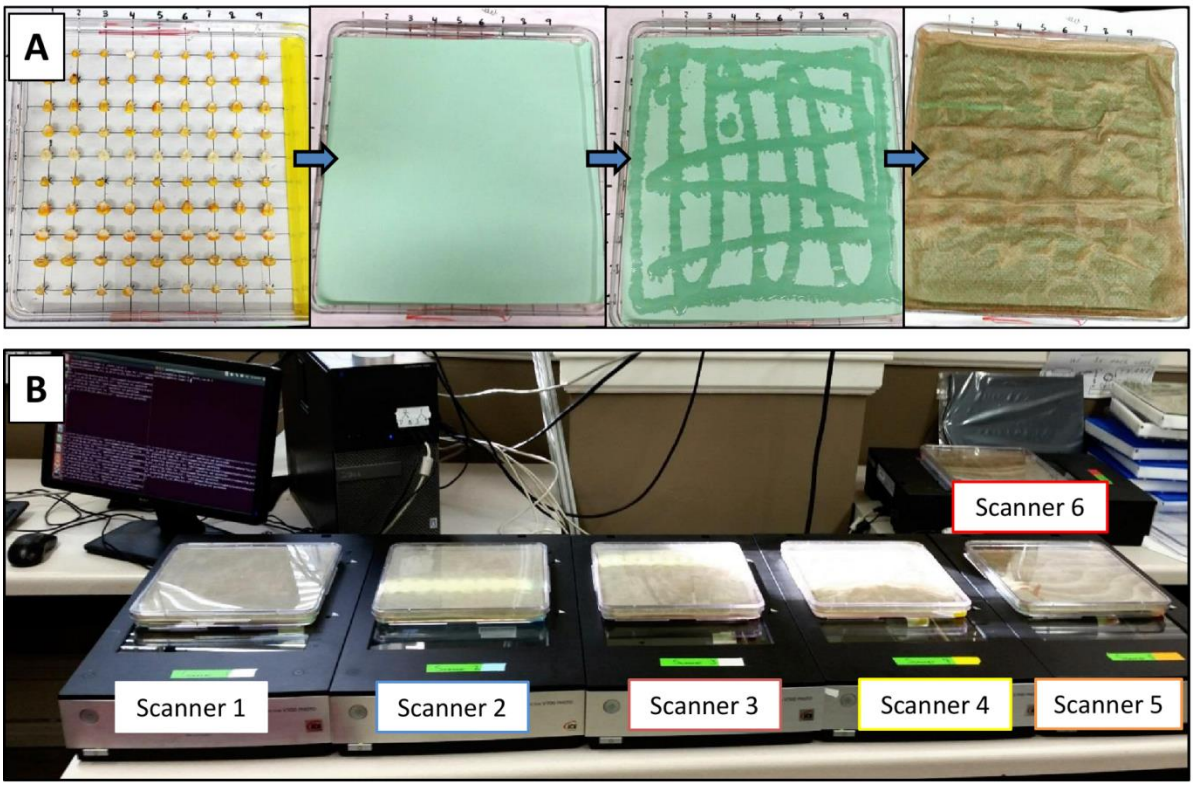


Figure 1. Maize seed imbibition phenotyping platform set-up. (A) Uniform, flat maize kernels were aligned with a square grid template, with one inch spacing between kernels, and were placed on square bioassay dishes filled with 150 mL of 1% agar per petri dish. Nine kernels per genotype (occupying one column each) per row were placed on the agar petri dishes, with ten genotypes placed per agar petri dish. Upon completion of kernel plating, kernels were covered with one layer of green paper to serve as background contrast for image analysis, and were subsequently covered with two layers of paper towel, evenly soaked with 20 mLs of deionized H_20 to saturate the paper towels to imbibe the kernels. (B) Upon wetting, agar plates were immediately transferred to flat-bed scanners (Epson Perfection V700 photo flatbed scanner) for image acquisition and analysis (one petri dish per scanner, six scanners running per day). Kernels were imbibed at room temperature (20ºC).

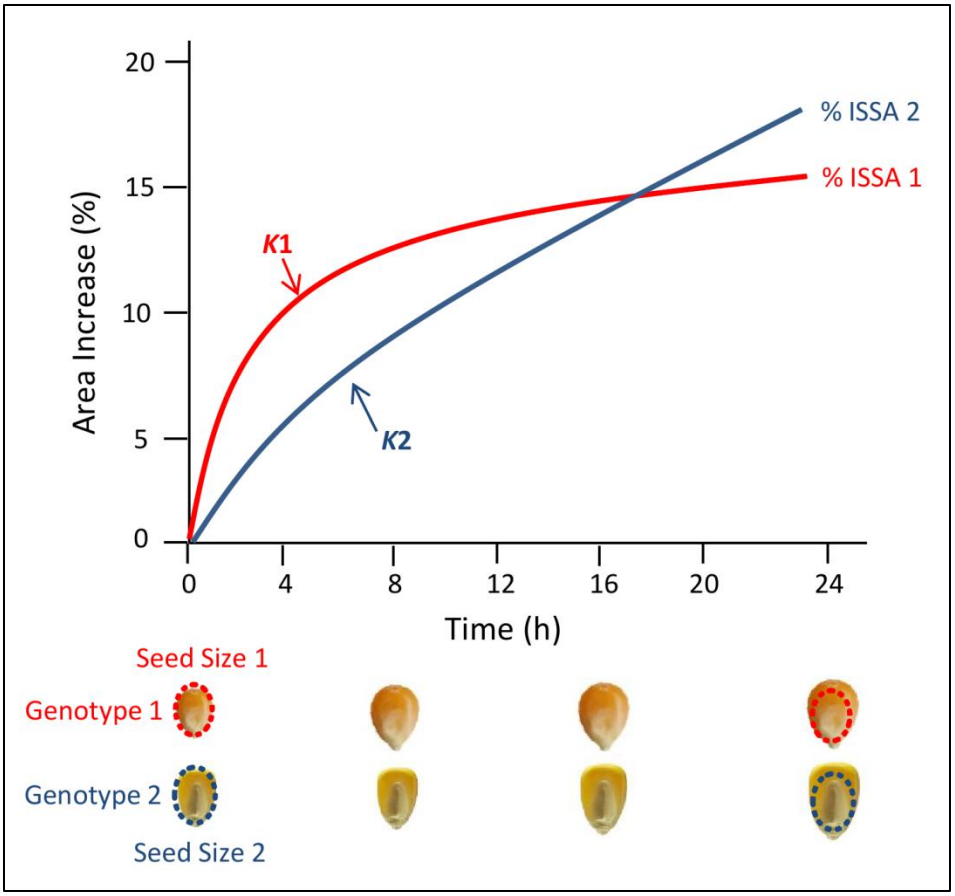


Figure 2. Example schematic depicting maize kernel imbibition phenotypes extracted from image analysis for two contrasting genotypes. Maize kernel images were derived from flatbed scanner images taken every ten minutes for 24 H, and were exported to MATLAB to derive three different kernel phenotypes: initial seed size (measured in pixels), percent increase in seed surface area after 24 hours (%ISSA), and imbibition rate, given as a constant termed "*K*." %ISSA was calculated using the formula $\%ISSA = (Area_{24Hr}-Area_{initial})/Area_{initial.}$ Imbibition rate constant *K* was calculated by fitting the raw kernel area at every time point over 24 hours to a previously defined equation: $Area_{max}(1 - e^{-Kt}) = Area(t)$, solving for *K* (LEOPOLD 1983). Example phenotypes are colored red and blue for genotypes one and two, respectively.

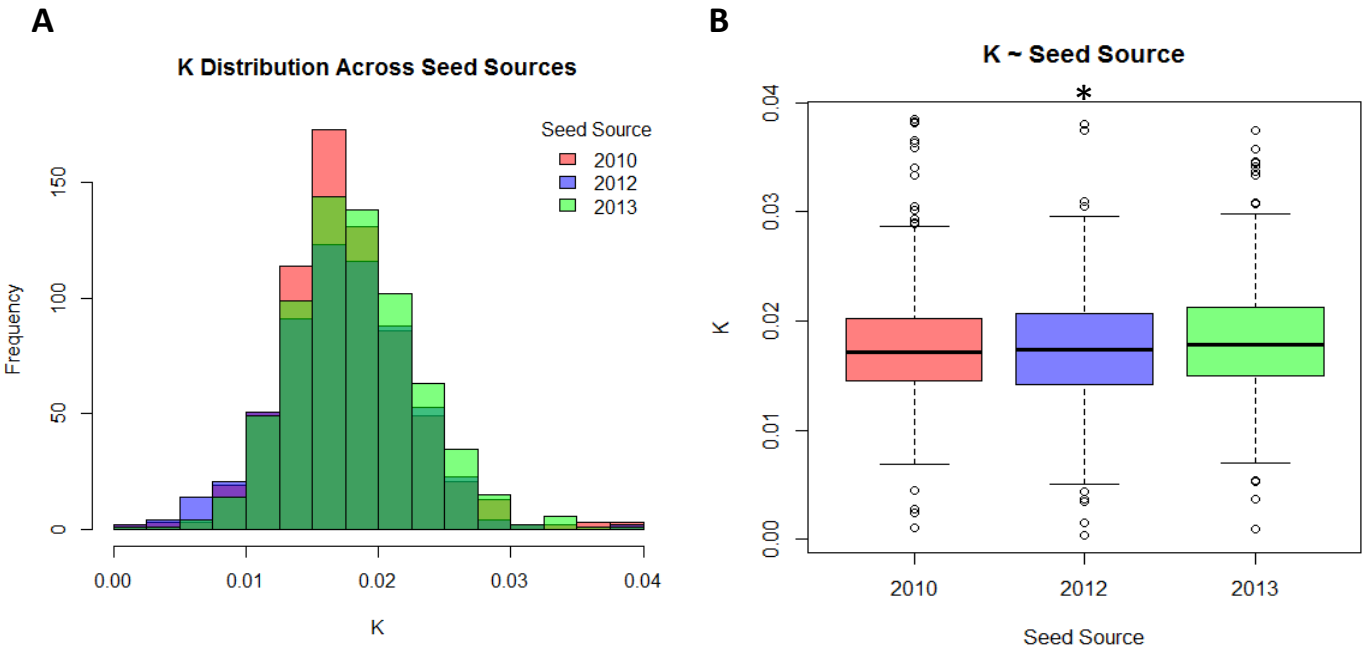


Figure 3. Effect of seed source on the phenotypic distribution of imbibition rate constant (*K*) for 25 diverse inbreds sampled from the WiDiv population. (A) Super-imposed histograms depicting phenotypic distributions of trait *K* from seeds derived from three separate growing years (2010, 2012, and 2013). (B) Box-plots depicting the means and distributions of *K* by seed source year. *$p < 0.05$

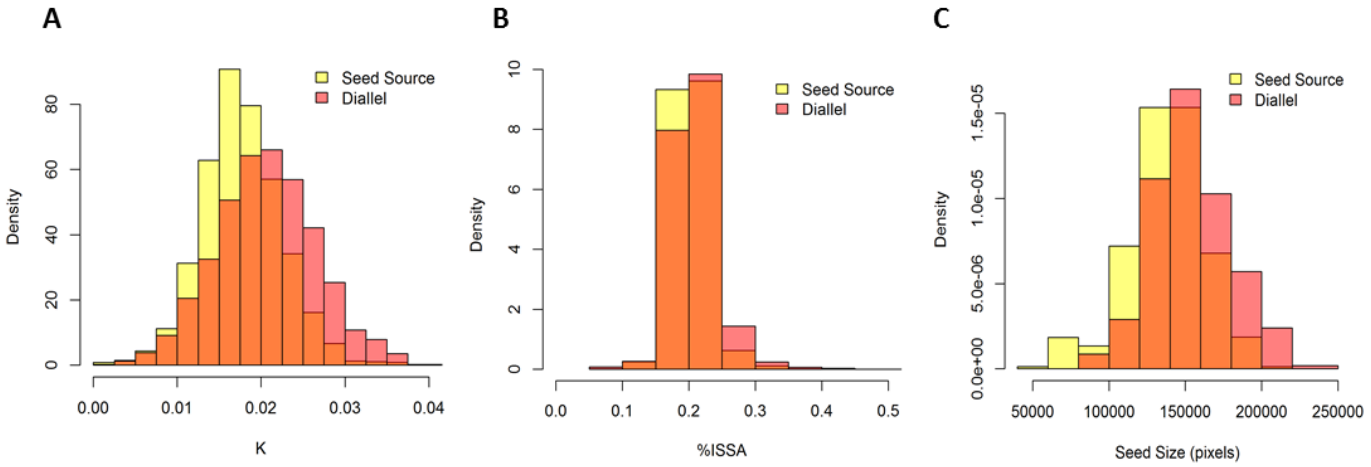


Figure 4. Histograms depicting phenotypic variation for three measured traits (*K*, %ISSA, and seed size) observed in both the seed source and diallel experiments. (A) Histogram of *K*, the imbibition rate constant from both the seed source and diallel experiments. (B) Histogram of %ISSA, the percent increase in seed surface area after 24H, from both the seed source and diallel experiments. (C) Histogram of seed size (measured in pixels) from both the seed source and diallel experiments.

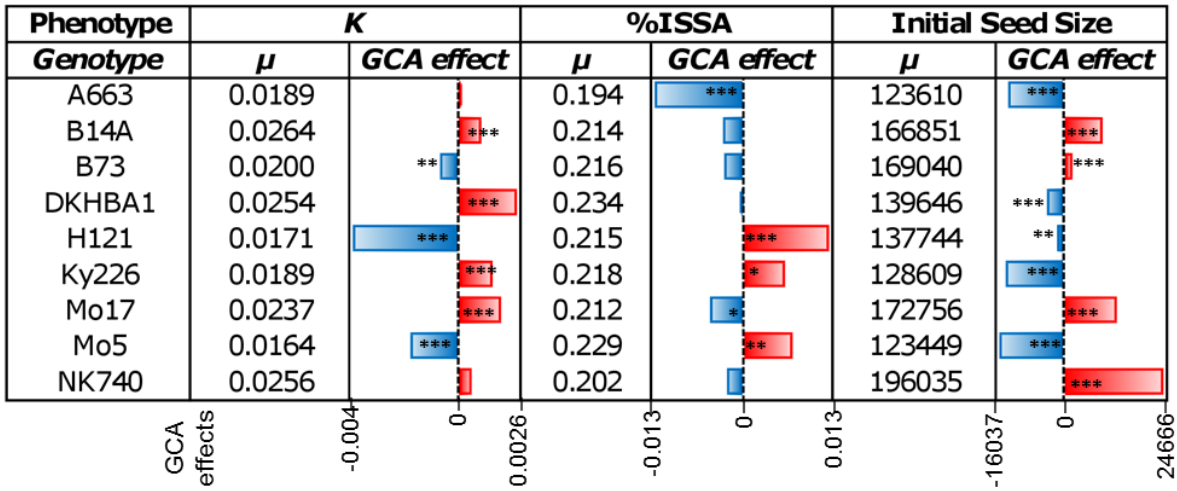


Figure 5. General combining ability (GCA) and mean phenotypic values of maize imbibition traits of the nine parents used in the full-diallel. Maize kernel imbibition traits measured include initial seed size (measured in pixels), percent increase in seed surface area after 24 hours (%ISSA), and imbibition rate, termed *K*. Red and blue bars indicate positive and negative GCA effects, respectively. GCA effects which are statistically significantly different from zero as calculated by *F*-test are given by ***$p < 0.001$, **$p < 0.01$, and *$p < 0.05$. Non-significant values are not marked with an asterisk.

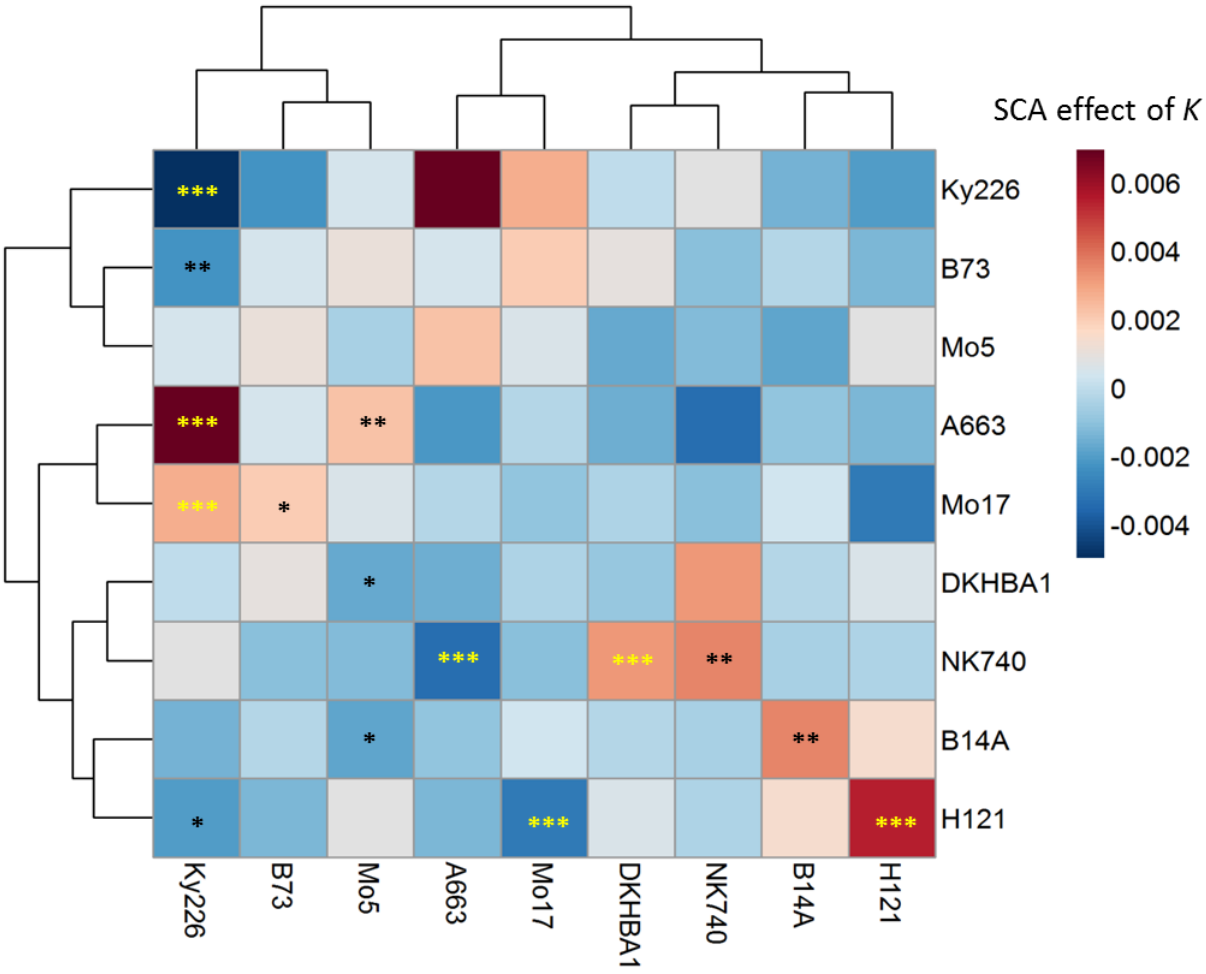


Figure 6. Heat map of the specific combining ability (SCA) effects matrix of imbibition rate *K* for thirty-six hybrid combinations used in the full-diallel experiment. The SCA effect matrix shown above is symmetrical, with significant values denoted in the lower half of the diagonal. SCA effects which are statistically significantly different from zero as calculated by *F*-test are given by ***$p < 0.001$ (colored yellow), **$p < 0.01$, and *$p < 0.05$. Non-significant values are not marked with an asterisk.

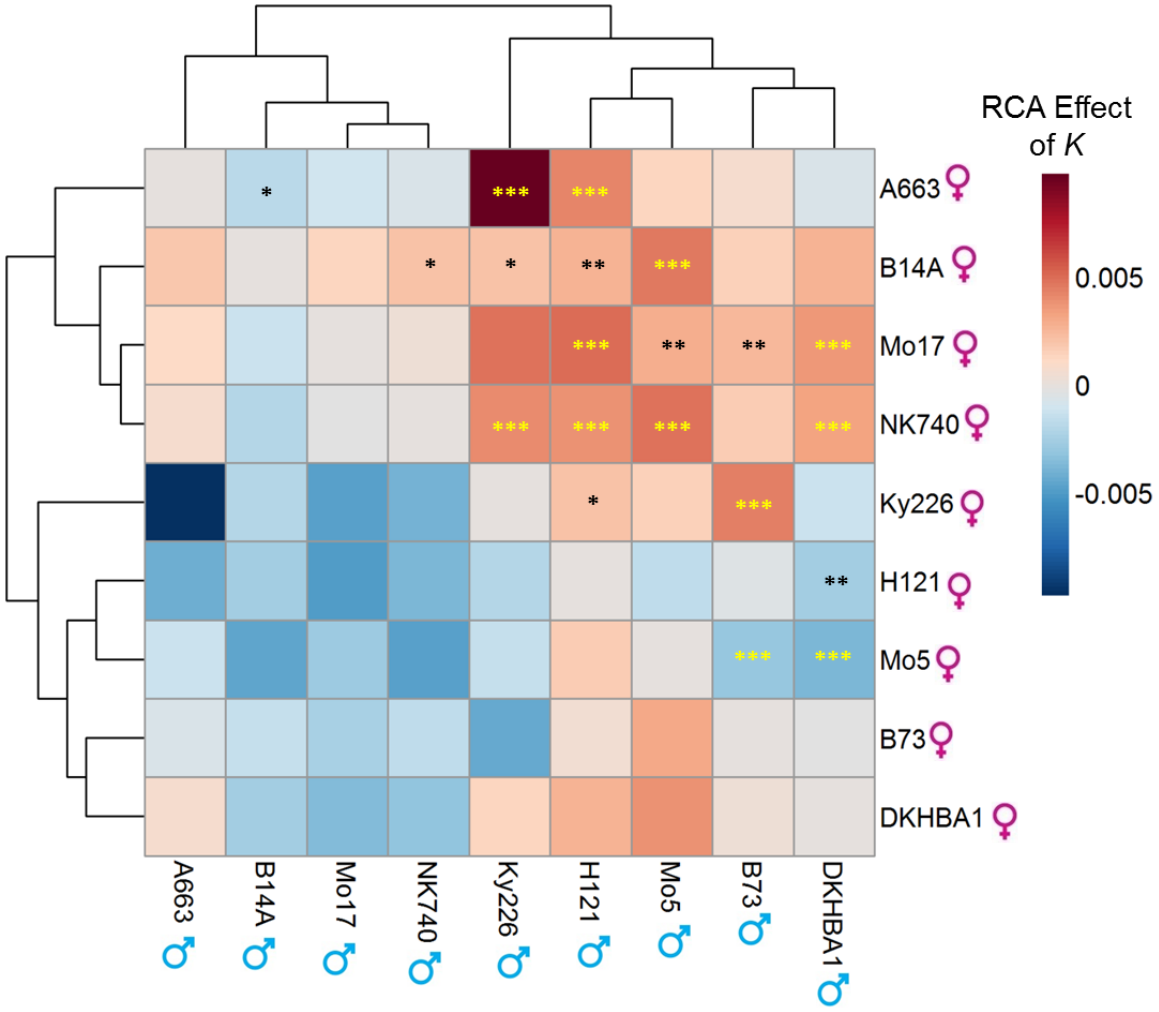


Figure 7. Heat map of the reciprocal combining ability (RCA) effects matrix of imbibition rate *K* for 72reciprocal hybrid combinations used in the full-diallel experiment. The RCA effect matrix is symmetrical, with values inverted depending on the direction of the cross, with males and females marked as shown. Significant values are denoted only in the upper-half of the diagonal of the matrix. RCA effects which are statistically significantly different from zero as calculated by *F*-test are given by ***$p < 0.001$ (colored yellow), **$p < 0.01$, and *$p < 0.05$. Non-significant values are not marked with an asterisk.

Acknowledgement of co-authorship for Chapter 3

Authors: S. Stelpflug[1], N. Miller[2], E. Spalding[2], N. de Leon[1] and S.M. Kaeppler[1]

Affiliations: [1]Department of Agronomy, University of Wisconsin, 1575 Linden Drive, Madison, WI 53706; [2]Department of Botany, University of Wisconsin, 132 Birge Hall, 430 Lincoln Drive, Madison, WI 53706

Author's contributions: SS, NM, ES, NDL and SK conceived and designed the study. SS collected the data, performed statistical and genetic analyses, and wrote the manuscript. NM designed the phenotypic assay and the programming code to analyze and export phenotypic data. All authors read and approved the final manuscript.

CHAPTER 3. INTEGRATING "OMICS" DATA TO REVEAL GENOTYPE-PHENOTYPE ASSOCIATIONS UNDERLYING MAIZE SEED IMBIBITION TRAITS

ABSTRACT

Imbibition, defined as the rapid uptake of water by the dry seed, represents the first phase of germination and has been shown to affect emergence rate and stand establishment, all of which are important pre-requisites affecting optimum yield potential. To evaluate this trait, we utilized a semi-automated image-based phenotyping platform across time to measure imbibition rate *K* and percent increase in seed surface area 24 hours post-imbibing (%ISSA). Genome-wide association mapping was performed on both of these imbibition traits utilizing ~437K RNA-seq based markers across 500 diverse maize inbred lines, identifying 104 trait-associated SNPs in total. Using a systems biology-based multi-omics approach, we were able to resolve genetic regions associated with imbibition rate *K* and identified two candidate genes: *ZmAMY1A* (GRMZM2G103055), encoding a putative α-amylase protein, and GRMZM2G172043, a putative β-1,3-glucanase protein. SNP-based pathway enrichment analysis indicated over-representation of biological processes related to carbohydrate metabolism, hydrolase activity, protein degradation, translation, membrane structure, and kinases involved in signaling cascades. These terms overlapped with both our identified candidate genes and global processes known to be differentially expressed during imbibition and early germination in other plant species. We also demonstrated a highly significant relationship of imbibition rate *K* with radicle emergence time and total germination percentage, indicating that our rapid phenotyping platform exhibits potential agronomic utility. Overall, this study has helped characterize the genetic architecture of seed imbibition in maize and has demonstrated the power of integrating multi-omics datasets to better identify candidate genes underlying genotype-phenotype relationships.

3.1 INTRODUCTION

Understanding the genetic basis of complex traits remains a perpetual quest for researchers across disciplines and species. Recent technological advances in data generation from multiple levels of biological systems — including next-generation DNA sequencing (METZKER 2010), RNA expression levels (WANG *et al.* 2009; OZSOLAK AND MILOS 2011), methylation patterns (LAIRD 2010), chromatin binding sites (PARK 2009), proteomics (ALTELAAR *et al.* 2013), post-translational modifications (SCHWÄMMLE *et al.* 2015), and metabolomics (SHULAEV 2006) — have driven the field of translational genomics for the past decade, producing ever-increasing amounts of useful data. Historically, each type of data has been considered independently to examine relationships with biological processes, and using these methods, researchers have assembled some of the puzzle pieces underlying complex-trait genetic architectures. However, much of the genetic etiology of quantitative traits and biological networks remains unexplained, which could be partially due to the focus on restrictive, singular data sets. To combat this problem, a systems-based genomics approach can achieve a more thorough and revealing interrogation of genotype-phenotype associations than an analysis using a sole data type (HAWKINS *et al.* 2010). Combining multiple forms of 'omics' data can compensate for absent or unreliable information in any one data set, and multiple sources of evidence pointing to the same gene or pathway are less likely to lead to false positives (RITCHIE *et al.* 2015). Importantly, a more comprehensive biological model is only likely to be revealed if the different levels of genetic, transcriptional and proteomic regulation are considered in an analysis.

The utilization of a systems-based genomics approach is particularly crucial for the understanding of genetic architectures and networks of seed traits in maize, owing to the lack of

correspondence between transcriptional, proteomic, and metabolome levels attributable to dynamic post-transcriptional and post-translational regulation in these tissues (RAJJOU *et al.* 2012; WALLEY *et al.* 2013; GALLAND *et al.* 2014). Additionally, maize seed phenotypes are influenced by seed parent effects due to differential genetic dosages in unique seed compartments and/or parental imprinting of alleles (ROACH AND WULFF 1987; GUTTERMAN 2000; WATERS *et al.* 2011), further adding to the genetic complexity. For instance, seed germination and seedling establishment are traits of economic and ecological importance as they determine the effectiveness of the initial introduction of plants into the environment. In maize breeding, rapid germination rate and overall emergence is a key pre-requisite for obtaining satisfactory stands, seedling establishment, and seedling vigor, all of which ultimately affect optimum yield potential (SHAW AND LOOMIS 1950; MARTIN *et al.* 1988).

Germination *per se* commences with the rapid uptake of water by the quiescent seed, initiating an orderly transition of increased hydration, enzyme activation, storage product breakdown, cellular metabolism, elongation of the embryonic axis, and ends with the penetration of the structures surrounding the embryo by the radicle (BEWLEY 1997). Water uptake during seed germination can be described by a classic triphasic model (BEWLEY 1997). Phase I of this process, imbibition, occurs due to the rapid absorption of water by cell-wall and protoplasmic macromolecules contained within the seed, such as proteins and polysaccharides, wherein water molecules are held by electrostatic forces such as hydrogen bonds (NOGGLE AND FRITZ 1983). During imbibition, the seed rapidly swells and changes in size and shape (ROBERT *et al.* 2008; PRESTON *et al.* 2009). This movement of water into the seed is due to diffusion and capillary action with water moving from a region of higher water potential (the environment) to the seed. Dry maize seeds have extremely low water potentials (as low as -400 megapascals) attributed to

their osmotic and matric characteristics (SHAYKEWICH 1973; VERTUCCI 1989). Uptake of water during imbibition is essential to the initiation of cellular metabolism for at least three reasons: to activate enzymes, to solubilize and transport reactants, and to serve as a reactant itself, especially in the hydrolytic digestion of stored reserves of proteins, carbohydrates, and lipids (WOODSTOCK 1988). Phase II is often referred to as the plateau phase, as increases in fresh weight due to physical swelling are leveling off. The plateau phase involves a cascade of metabolic processes initiated in the seed, such as starch hydrolysis, protein degradation, mitochondrial assembly, and glycolysis, leading to cell differentiation and elongation. Per definition germination is completed upon entry into Phase III, in which growth becomes visible and the primary root ruptures the covering seed layers to complete germination (SCHOPFER AND PLACHY 1984; BEWLEY AND BLACK 1994). This process involves a complex "give and take" feedback loop between the expansive force of the embryo and the mechanical restraining force imposed by the endosperm and testa. In all seeds in which the embryo must penetrate the surrounding structures, as is the case for maize, embryo growth potential must exceed the mechanical resistance of these tissues for radicle emergence to occur (LIPTAY AND SCHOPFER 1983; BRADFORD 1986).

To breed vigorous plant cultivars with fast and uniform field emergence, it is crucial to understand the genetic factors that contribute to an adequate germination performance and seedling growth. A common method for identification of genomic regions associated with complex quantitative traits is genome-wide association (GWAS) mapping, in which marker-trait associations are calculated across a broad set of diverse germplasm in order to define chromosome regions harboring promising genes (INGVARSSON AND STREET 2011). The attraction of using a GWAS approach in maize is the ability to resolve associations to single genes due to

thousands of captured ancient meioses in a set of unrelated individuals and the rapid breakdown of linkage disequilibrium (LD) which exists in highly inter-mated, diverse maize populations (REMINGTON *et al.* 2001; TENAILLON *et al.* 2001). However, a common drawback of GWAS is that population structure can result in non-random distribution of alleles within a diversity panel. This can lead to the detection of false-positive marker associations (KORTE AND FARLOW 2013) or a non-detection of rare or sub-population specific alleles (NORDBORG AND WEIGEL 2008). In such a case, statistical methods have to be applied which correct for population stratification. As mentioned previously, seed germination is a very complex trait controlled at the transcriptional, translational, and metabolic level (RAJJOU *et al.* 2012), adding to the difficulties of identifying contributing genetic factors by conventional genetic GWAS. However, by combining these omics datasets across both levels of the central dogma and across species, researchers can leverage this information to sort through both the false-positive and the linked associations identified with singular GWAS, and instead look at all levels of regulation to identify and characterize true variants, which will aid our understanding of genetic architectures going forward.

In short, the primary objectives of this study were to 1) use semi-automated phenotyping platform to phenotypically characterize maize seed imbibition rate over time and total amount of seed water uptake in inbred lines of the Wisconsin Diverse (WiDiv) association panel, 2) use both GWAS and pathway enrichment analysis to identify both global and local gene candidates responsible for the genetic architecture of early maize imbibition and germination and 3) to integrate pre-existing syntenic, transcriptomic, and proteomic data to further resolve putative gene candidates identified via GWAS.

3.1 MATERIALS AND METHODS

Genetic Materials

Seed from 500 maize inbred lines from the Wisconsin Diverse (WiDiv) association panel was used for the GWAS study (HANSEY *et al.* 2011; HIRSCH *et al.* 2014) (**Appendix A4**). Seeds were not produced from a uniform seed source due to the fact that inbred increases spanned multiple summer nurseries (2008-2014) at the West Madison Agricultural Research Center (Madison, WI). A subset of 30 lines identified based on phenotype (**Appendix A5**) were utilized for a more detailed analysis of the relationship of seed imbibition rate and later germination traits. Genotypic data for the WiDiv association panel was generated by calling 436,576 high-quality single nucleotide polymorphisms (SNPs) using next-generation RNA-sequencing (RNA-seq) of two-week-old, whole plant seedling tissue, as described previously (HIRSCH *et al.* 2014). SNPs were then filtered to retain only di-allelic SNPs, and were imputed using the population-based haplotype clustering algorithm fastPHASE version 1.4.0, using default settings except for having a fixed number of 20 clusters (i.e. K = 20) (HOWIE *et al.* 2009; HIRSCH *et al.* 2014).

Phenotypic Analysis

Phenotypic analysis of the 500 inbred lines was conducted using a randomized complete block design (RCBD), blocked in time via two technical runs or replications sampling nine uniform kernels per genotype per replication. Phenotypic measurements were derived from imaged-based analysis for three different maize kernel traits (assay and method described previously in "Materials and Methods" of Chapter 2): initial seed size (measured in pixels), percent increase in seed surface area after 24 hours (%ISSA), and imbibition rate, given as a constant termed "*K*." %ISSA was calculated using the formula %ISSA = ($Area_{24Hr}$-

$Area_{initial}$)/$Area_{initial.}$ Imbibition rate constant *K* was calculated by fitting the raw kernel area at every time point to a previously defined equation: $Area_{max}(1 - e^{-Kt}) = Area(t)$, solving for *K* (LEOPOLD 1983), and was multiplied by a constant of 1000 for ease of interpretation.

In order to test the putative effect of imbibition rate *K* on germination characteristics, we selected 30 genotypes of the total 500 screened lines from the WiDiv population which displayed the most variation for *K*, sampling the ten fastest, slowest, and ten most intermediate genotypes from the distribution. The phenotypic assay used to score both percent total germination (%TG) and time to radicle emergence (TRE), given in hours, mirrored the agar plate scanner assay used to extract imbibition traits, as described previously (Chapter 2). Modifications to the kernel imbibition assay for the purpose of screening germination traits included: 1) images were acquired on Epson V700 flat-bed scanners every 30 min as opposed to 10 min over the course of 72 hours as opposed to 24 hours for the imbibition assays, and 2) five genotypes were used per agar plate as opposed to ten, alternating every other row in order to facilitate space for root growth. The germination experiment was conducted using two replications of nine kernels per genotype, blocked by runs of time. TRE was manually scored by examining sequences of images over time (0-72 H), and %TG was calculated by dividing the number kernels per genotype with emerged radicles after 72 H by the total number of kernels used.

Phenotypic data analysis

Phenotypic data from the WiDiv was analyzed using the *lmer* function of the *lme4* package implemented in R (R DEVELOPMENT CORE TEAM 2013; BATES *et al.* 2015) using the following mixed linear model:

$$[1]\ Y_{ijk} = \mu + G_i + B_j + \varepsilon_{ij},$$

where Y_{ijk} was the response variable of the *i*th genotype (*G*) of the *j*th replication blocked in time (*B*). The residual error ε_{ij} was assumed to be independent and following a normal distribution (~iidN(0,σ_{ε}^2)). Genotype and error were considered random and fixed effects, whereas replication was considered to be a fixed effect. Best linear unbiased predictions (BLUPs) were calculated and used as phenotypic inputs for the genome-wide association (GWAS) study. Variance components for the random effects of genotype and error were estimated with the restricted maximum likelihood method (REML). Using these estimated values, repeatability (R) was calculated for each trait measured using the following formula:

$$[2] = \frac{\sigma_G^2}{\sigma_G^2 + \sigma_e^2/r},$$

where σ_G^2 is the genotypic variance, σ_e^2 is the error variance, and r is the number of technical replications blocked in time. In order to determine the similarity of phenotypic values between technical replications, the *cor.test* function implemented in the base package of R (R DEVELOPMENT CORE TEAM 2013) was used to calculate Spearman rank correlations between the two replications for each trait measured. Pair-wise Pearson's product moment correlations were also calculated to determine the relationships between all of the phenotypic values measured across the WiDiv diversity panel using the *cor.test* function of R (R DEVELOPMENT CORE TEAM 2013).

In the selected subset of 30 lines (listed in **Appendix A5**), we assessed the relationships between imbibition rate *K*, %TG, and TRE via two tests: 1) pairwise t-tests between three divergent *K* groups, classified as "fast," "intermediate," and "slow" as mentioned above with the *t.test* function embedded in the core R program and 2) Pearson's correlations (and their

significance levels) among phenotypic means using the *cor.test* function in the core R program (R DEVELOPMENT CORE TEAM 2013).

Genome-wide association analysis

GWAS were performed for all traits using BLUPs estimated from equation [1] as input trait variables, testing the effects of 436,576 RNA-seq SNPs across the 500 genotypes using the GAPIT package (LIPKA *et al.* 2012) implemented in the R software program (R DEVELOPMENT CORE TEAM 2013). To accurately test the effect of individual SNP markers, we accounted for both polygenic background and population structure effects using the "Q + K" mixed-linear model approach (YU *et al.* 2006) given as follows:

$$[3]\ \mathbf{y} = \mathbf{Wm} + \mathbf{Qv} + \mathbf{Zu} + \mathbf{e},$$

where **y** is a vector of phenotypic BLUPs estimated from the mixed linear model above [1], **m** is a vector of SNP effects, **v** is a vector of population structure group effects, **u** is a vector of random polygene background effects, and **e** is a vector of residual effects. **W** and **Z** are incidence matrices of ones and zeros relating **y** to **m** and **u**, respectively (YU *et al.* 2006). The kinship matrix "**K**" (referred to as **Z** in the above mixed linear model) used to estimate polygene background effects was estimated with the VanRaden method (VANRADEN 2008) using the GAPIT package (LIPKA *et al.* 2012). This **K** matrix estimates the variance associated with **u**, which was assumed to be equal to $2\mathbf{K}\sigma_g^2$. The "**Q**" matrix was also estimated within the GAPIT package using principal component analysis (PCA) of the genotypic matrix (LIPKA *et al.* 2012). For both traits, we used stepwise regression as assessed by the Bayesian information criterion (BIC) and expected versus observed quantile-quantile (Q-Q) plots to determine model

assessment and relative fit, which in turn determined the number of PCs to include in each GWAS.

In order to determine an appropriate genome-wide threshold of significance which accounted for experiment-wise type-I error rate, we used the simpleM method (GAO *et al.* 2008; GAO *et al.* 2010; GAO 2011). By estimating linkage disequilibrium (LD), the non-random association of different alleles, between each pair of markers in the RNA-seq SNP dataset and by applying PCA to obtain the eigenvalues, the effective number of independent tests (N) can be determined for which a stringent Bonferroni correction can be used. In order to determine the number of independent tests, the number of tests was equal to the number of eigenvalues required to explain 99% of the explained variance (GAO *et al.* 2008; GAO *et al.* 2010; GAO 2011). For this study, the effective number of independent tests was N=204,513, thus the genome-wide threshold was equal to $p = 2.44\text{x}10^{-7}$. However, although Bonferroni-based multiple-testing corrections help to minimize type-I error (α, the detection of false associations), the lower α value results in a proportional increase in type-II error (β, the rejection of true associations), which reduces power to detect true trait-associated SNPs (TASs), indicating Bonferroni corrections may be too conservative (STREINER AND NORMAN 2011). Hence, in addition to the simple*M*-based Bonferroni (α=0.05/N) genome-wide threshold, we also displayed a comparatively more liberal "suggestive" genome-wide threshold equal to 1/N, which was equal to $p = 4.90\text{x}10^{-6}$, as has been performed in other GWA studies in plants (YANG *et al.* 2014).

In order to investigate relationships between gene models and identified significant SNPs, LD heatmaps were generated using the *LDheatmap* package implemented in R (SHIN *et al.* 2006; R DEVELOPMENT CORE TEAM 2013). LD between two loci was calculated as $r^2 = \frac{D^2}{p_1\, q_1\, p_2\, q_2}$,

where D is the deviation of the observed versus expected haplotype frequencies and p and q are the respective allele frequencies for the two loci (FALCONER AND MACKAY 1996).

Pathway-based association analysis in the WiDiv association panel

Pathway-based enrichment analysis was performed for the imbibition rate (*K*) measured in the WiDiv panel, as this trait was of the most biological importance and significant. Gene models having at least one significant SNP (based on a nominal *p*-value ≤ 0.005 from the GWAS of imbibition rate *K*) were included in the input gene lists used for analysis, excluding redundant gene models. The Plant Gene Set Enrichment Analysis toolkit (PlantGSEA) (YI *et al.* 2013) was used to perform a Singular Enrichment Analysis (SEA), defined as the over-representation of significant biological terms in our input gene list relative to the frequency of these biological terms in the whole-genome background gene list (*n*=23,900, corresponding to the number of genes having genotyped RNA-seq SNPs in the WiDiv association panel) (HIRSCH *et al.* 2014). This enrichment tool interrogates gene sets from three databases: (1), the Gene Ontology (GO) gene set, encompassing three overarching categories: biological processes (BP), molecular functions (MF), or cellular components (CC) (http://www.geneontology.org) (ASHBURNER *et al.* 2000); (2) Gene family-based gene sets, and (3) Publically curated gene sets including PlantCyc (ZHANG *et al.* 2010), the Kyoto Encyclopedia for Genes and Genomes (KEGG) (KANEHISA AND GOTO 2000; KANEHISA *et al.* 2004), and Plant Ontology (PO) (AVRAHAM *et al.* 2008). A test of proportions based on the cumulative hypergeometric distortion of Fisher's exact test was used to assess the association of a given biological pathway with the trait examined (UPTON 1992; TAVAZOIE *et al.* 1999). To account for multiple testing, we controlled for the false discovery rate or FDR (BENJAMINI AND HOCHBERG 1995) and declared a pathway to be significant if its *q*-value (FDR adjusted *p*-value) was ≤ 0.05.

3.3 RESULTS AND DISCUSSION

Phenotypic analysis of maize seed and imbibition traits in the WiDiv association panel

Phenotypic distributions for each seed imbibition trait measured in the WiDiv association panel were examined and described (**Figure 1**). Trait variability for seed imbibition phenotypes, given by the mean and standard deviation, ranged from 18.91 ± 4.93 for imbibition rate constant *K*, 18.8% ± 3.7 for percentage increase in seed surface area (%ISSA), and 126,864 ± 21,637 pixels for seed size, indicating large levels of phenotypic variation for these traits. Overall, the mean and standard deviation values for all measured traits in the WiDiv were quite similar to the values reported for both the seed source and full-diallel experiments previously described (Chapter 2).

The relationship between seed size, *K*, and %ISSA was determined by Pearson's correlation analysis (**Appendix A6**). The only trait correlation that was highly significant was between *K* and %ISSA (r = -0.218, $p < 0.001$), which was similar in both magnitude and significance to the seed source and diallel experiments discussed previous (Chapter 2). This observation further confirms that seeds which reach a larger total surface area after 24H of imbibition tend to have a slower initial imbibition rate *K*, which is not unexpected given that %ISSA is used in the mathematical calculation used to derive *K* (LEOPOLD 1983). Additionally, imbibition rate *K* exhibited a minor, but significantly negative correlation to seed size in the WiDiv panel (r = -0.076, $p < 0.05$), which was weaker than the relationship observed in the seed source and diallel experiments (Chapter 2 and **Appendix A6**). Additionally, our results indicate a substantially weaker to non-existent negative relationship between *K* and seed size compared to the significantly negative correlation observed in *Medicago truncatula* (r = -0.65, $p < 0.001$)

(DIAS *et al.* 2011), although imbibition was measured less frequently and in terms of water weight in this study. Overall, the lack of correlation between seed and imbibition traits measured in a genetically diverse population such as the WiDiv suggests non-pleiotropic genetic control of these traits.

After fitting the phenotypic data with a mixed linear model [1], all seed and imbibition traits exhibited substantial genotypic variance within the WiDiv population ($p < 0.001$; **Table 1**). Repeatabilities for all traits measured were also very high (ranging from $R = 81.8\%$ for %ISSA to $R = 93.6\%$ for seed size; **Table 1**). Both the genotypic variances and repeatability estimates associated with the traits assayed from the WiDiv association panel exceeded the values reported for *K* and %ISSA in the seed source and diallel experiments described previously (Chapter 2), owing to the substantially larger population size ($n = 500$ genotypes for the WiDiv panel, $n = 81$ and $n = 25$ for the diallel and seed source experiments, respectively) and genetic diversity captured within the WiDiv panel. Accordingly, Spearman rank correlations (ρ) of genotypic means between the two replication runs blocked in time were also very high and statistically significant, with $\rho = 0.68$ for %ISSA, $\rho = 0.75$ for imbibition rate *K*, and $\rho = 0.86$ for seed size ($p < 0.0001$). Given the high repeatability values and Spearman rank correlations between replicates, BLUPs for each of the traits (*K*, %ISSA, and seed size) were calculated using both replications and were subsequently used for GWAS.

GWAS of maize seed and imbibition traits in the WiDiv association panel

GWAS were performed for seed imbibition rate *K* (**Figure 2**) and for %ISSA (**Appendix A7A**) for 500 maize inbred lines sampled from the WiDiv association panel. Using a Bonferroni correction based on the effective number of independent SNP markers (GAO *et al.* 2008; GAO 2011), the *p*-value thresholds were set at 4.90×10^{-6} (suggestive) and 2.44×10^{-7} (significant) as described previously (Materials and Methods). Step-wise model selection of the "**Q** + **K**" mixed-linear model and surveying of the expected versus observed quantile-quantile (Q-Q) plots for each trait determined that imbibition rate *K* did not require the additional fixed polygenic background effect (**Q**). For these traits, GWAS models accounting for familial relatedness through the VanRaden kinship matrix alone (i.e. the **K** model) adequately controlled for false positives without overfitting (**Appendix A8**). Alternatively, after examining the naïve model (**Appendix A7B**) for %ISSA, it was determined that this trait was highly influenced by population structure and required an inclusion of a **Q** matrix with 6 PCs in the model (**Appendix A7C**).

In this study, a total of 104 trait-associative SNPs (encompassing 67 genes) exceeding the genome-wide suggestive threshold were identified across all three traits, of which 22 associations encompassing 16 gene models exceeded the genome-wide significance threshold (**Table 2**). However, the majority of the suggestive TASs (100 of 104) and all of the significant TASs (22) were identified in the GWAS of trait %ISSA, which was somewhat surprising given the narrow phenotypic distribution of trait in the WiDiv panel (**Figure 1**); given the abundance of TASs identified in the genome-wide scan for trait %ISSA, only the genome-wide significant TASs for this trait are shown (**Table 2)**. Only four suggestive TAS contained within three genes were identified for imbibition rate *K* (**Table 2**). The majority of the TASs found in our study

explained less than 5% of the observed phenotypic variance, suggesting that a complex genetic architecture underlies these quantitative traits. One advantage of utilizing genomic data sets derived from RNA-seq is that TASs are enriched and found almost exclusively in genic regions, specifically on expressed exons for which we have information about the underlying gene models. Alternatively, as our RNA-seq SNPs were derived from mRNAs of whole seedling tissues (excluding seeds), this SNP dataset may be missing information for a significant number of important, seed-specific genes related to germination-like phenotypes.

Due to the fact that imbibition rate *K* represents Phase I of the triphasic physiological seed germination process and may be a component or predictive trait of this agronomically significant phenotype (SCHOPFER AND PLACHY 1984; BEWLEY AND BLACK 1994), all subsequent discussions and analyses of putative candidate genes and their biological functions will focus on TASs of imbibition rate *K*.

Analysis of putative candidate genes associated with imbibition rate K using a biological systems multi-omics approach

We utilized a biological systems multi-omics approach, combining local LD analysis of TASs with syntenic, transcriptomic, and proteomic analysis of our candidate genes using multiple publically available datasets (HOWELL *et al.* 2009; SCHNABLE *et al.* 2012; WALLEY *et al.* 2013; STELPFLUG *et al.* 2015) to identify and prioritize candidate genes. As stated previously, we identified four trait-associated SNPs belonging to three gene models which were significant at the suggestive genome-wide threshold for imbibition rate *K* (**Table 2** and **Figure 2**). The most significant TAS for trait *K* (rna5_209899708, $P = 8.61 \times 10^{-7}$), which was also quite close to the

genome-wide significance threshold, was located within gene model GRMZM2G103055 on chromosome 5; this TAS was in almost complete LD (R^2= 0.98) with the next most significant TAS, rna5_209899702, which is not unexpected given that these two SNPs are only six base pairs apart (**Figure 3A**). Long-range LD analysis over the physical distance (37kb) of this genetic region indicates low LD of this TAS ($R^2 < 0.25$) with all neighboring genic SNPs; additionally, examination of gene annotation function and RNA-seq gene expression analysis over 80 tissues (Stelpflug *et al.* 2015) indicates that the three closest neighboring genes to our candidate gene are 1) a chlorophyll-binding protein expressed exclusively in leaves 2) an ethylene-responsive binding element expressed exclusively in roots, and 3) a ferredoxin-nitrate reductase protein which exhibits extremely low expression in seed tissues (data not shown), implicating gene GRMZM2G103055 as the most likely candidate gene (**Figure 3A**). The variant effect predictor (VEP) tool within Gramene was used to determine the effect and consequences of our discovered SNP variants on gene, transcript, regulatory, and protein sequences (e.g. stop gained, stop lost, splice sites, missense, nonsense, synonymous, frameshift, etc.) (McLaren *et al.* 2010; Monaco *et al.* 2013). Although TASs rna5_209899702 and rna5_209899709 were located within the second exon of gene GRMZM2G103055, both are synonymous mutations which do not change the amino acid sequence (Alanine → Alanine) and are predicted by VEP to have a low effect (**Appendix A9**). Hence, these two TASs are most likely not causal variants, but are linked to a proximal causal variant not assayed via RNA-seq.

Gene model GRMZM2G103055 encodes a putative α-amylase protein according to its syntenic ortholog in rice, *RAmy1A* (Os02g0765600/ LOC_Os02g52700), which shares 75% amino acid identity to maize gene GRMZM2G103055 (hereafter referred to as *ZmAmy1A*) (SCHNABLE *et al.* 2012). Alpha-amylases are the primary enzyme family responsible for the

hydrolysis and mobilization of starch granules during the germination of cereal grains and have been shown to play a critical role in the vigor of seedlings (AKAZAWA *et al.* 1991; JONES AND JACOBSEN 1991; KARRER *et al.* 1992; PERATA *et al.* 1993; YOUNG *et al.* 1997). This enzyme class is regulated transcriptionally by hormones, as it is induced by gibberellins (GAs) in the scutellar epithelium and inhibited by abscisic acid (ABA) (ZWAR AND HOOLEY 1986; SKRIVER *et al.* 1991; GUBLER *et al.* 1995; KANEKO *et al.* 2002). Furthermore, studies in rice and maize have shown that various carbohydrate compositions of seeds can influence α-amylase activity. For example, *Amy3D* α-amylase mRNA levels in rice scutella are reduced when sucrose, glucose, fructose or maltose accumulate in the tissue (KARRER AND RODRIGUEZ 1992). Additionally, researchers studying sweet corn have found that enhanced concentrations of sugar due to starch-deficient endosperm mutations, such as *sugary1* (*su1*), *sugary enhancer 1* (*se1*), and *shrunken2* (*sh2*), all substantially delay α-amylase transcription in sweet corn, and hence affect germination emergence rate and total percentage of germination (SANWO AND DEMASON 1992; SANWO AND DEMASON 1993; YOUNG *et al.* 1997). Ratios of enzyme activity to transcript abundance were not consistent across different endosperm mutations, with transcription of α-amylase being equivalent between *Su1* and *sh2*, however there was 12x the α-amylase activity in the *Su1* genotype which was due to a post-transcriptional "block" of α-amylase production in the aleurone layer of *sh2* kernels (YOUNG *et al.* 1997). These results support the hypothesis that α-amylase gene expression is regulated in a tissue-specific manner involving a feedback control mechanism in the embryo and an osmotic control mechanism in the aleurone (YU *et al.* 1996).

Multiple studies in both maize and rice have shown that α-amylase gene expression and activity exists at very low levels in the quiescent seed, but drastically increases during early germination (DURE 1960; O'NEIL *et al.* 1990). Indeed, both transcriptional and proteomic

analysis from two publically available datasets of *ZmAMY1A* in maize inbred B73 indicates low expression in the developing embryo, endosperm, pericarp and aleurone of the maize seed (WALLEY *et al.* 2013; STELPFLUG *et al.* 2015) (**Figure 3B and 3C**). However, *ZmAmy1A* displays a 4-5 fold increase in both transcript and protein expression after 24 and 48 hours post imbibition, respectively (**Figure 3B and 3C**). Previous *in situ* gene expression analysis of rice gene *RAmy1A* indicates that α-amylase transcripts are first detected at 6 hours after imbibition (6HAI) in both terminal regions of the scutellar epithelium of the embryo, which is secreted to the region of the aleurone layer closest to the embryo around 24 HAI (SUGIMOTO *et al.* 1998). *RAmy1A* gene expression has subsequently been confirmed to be upregulated even earlier in the imbibition process (3HAI) based on a RNA-seq time-course experiment of seed germination in rice (HOWELL *et al.* 2009). Analyses using embryo defective mutants of rice concluded that the embryo directly regulates *RAmy1A* gene expression in aleurone layer cells and that removal of the scutellum abolishes α-amylase gene expression globally (SUGIMOTO *et al.* 1998). Accordingly, it has been suggested the scutellum epithelium is indispensable for α-amylase gene expression in the aleurone, further demonstrating the complex cross-talk between seed tissues during early germination processes (SUGIMOTO *et al.* 1998). Based on these combined findings, α-amylase gene *ZmAMY1A* may be an appealing gene candidate to target for further molecular, genetic, and physiological characterization to test its putative effect on imbibition rate and potential impact on germination characteristics in maize.

An additional TAS (rna5_7241582) for imbibition rate *K* was found on chromosome 5, located within gene model GRMZM2G172043. Very low levels of LD are observed between this TAS and the entire set of neighboring genic SNPs ± one gene model over a 34.4 kb region (**Figure 4A**), implicating GRMZM2G172043 as a candidate gene deserving of more analysis.

The low amounts of LD typically observed in the WiDiv panel were not unexpected given that within diverse maize inbred lines, LD tends to decay to very low values ($r^2 < 0.1$) within 1.5 kb (REMINGTON *et al.* 2001) and has been shown to decay from 0.95 to 0.20 within 100-200 bp (TENAILLON *et al.* 2001). The rare allele (T, 0.088 frequency) has an allelic effect of 4.2 on *K* relative to the major allele "A" (**Appendix A10A**)*;* interestingly, VEP analysis of this allelic substitution indicates that this TAS is located within the two base-pair region splice-site of the first intron of gene GRMZM2G172043, and is predicted to have a large effect on the function of this gene (**Appendix A9** and **Appendix A10B**). Alteration of this splice site could result in the inclusion of the first intron, and hence a frameshift mutation in the translated protein, producing a null function. Furthermore, this candidate gene is syntenic and orthologous to rice locus Os03g53860 (Putative periplasmic β-glucosidase, 85% amino acid identity) and *Arabidopsis* locus At5g20950 (Exhydrolase II, 65% amino acid identity) (SCHNABLE *et al.* 2012), and is also predicted to encode a periplasmic β-glucosidase protein. 1,3;1,4-β-Glucans, while absent from dicots, are the predominant non-cellulosic polysaccharides found in the cell walls of monocot grasses (HØJ AND FINCHER 1995; HRMOVA AND FINCHER 2001); they are especially abundant in the starchy endosperm of cereal grains, where they account for up to 75% of the wall polysaccharides (LEUBNER-METZGER 2003). During germination, enzymes with β-glucanase activity hydrolyze β-glucans containing mixed β-1,3 and β-1,4 glycoside linkages and provide access for other hydrolytic enzymes during the post-germination mobilization of storage reserves, and may participate in the initial release of β-glucans from endosperm cell walls (BATHGATE *et al.* 1974; LEUBNER-METZGER 2003). Endo-type β-glucanases are present in the ungerminated grain and rise markedly during germination (SIMMONS 1994; HRMOVA AND FINCHER 2001). In dry, mature caryopses of barley and wheat, endo-type β-glucanase activity is

found to be most associated with the embryo (particularly the scutellum). However, this activity spread out homogeneously throughout the seed, increasing markedly in the aleurone layer and the starchy endosperm during germination; additionally, GA treatment of whole grains or isolated aleurone layers enhances both the timing and amount of secretion of β-glucanase (SIMMONS 1994; HØJ AND FINCHER 1995; HRMOVA AND FINCHER 2001; LEUBNER-METZGER 2003; VATANDOUST *et al.* 2012).

Intriguingly, expression of maize candidate gene GRMZM2G172043 at both the transcriptome and proteome levels indicates very high expression in the earliest stages of developing embryos, decreasing in a step-wise manner during seed maturation and grain filling (**Figure 4B** and **Appendix A11**). Due to the associations of candidate gene GRMZM2G172043 with the embryo, scutellum, and its secretion in the periplasm, this gene's most likely function is as an endo-type β-glucanase, potentially participating in cell-wall remodeling and cell-wall breakdown during early germination and imbibition. Two independent gene expression time-course experiments of imbibition in rice and *Arabidopsis* indicate that both candidate gene syntenic orthologs, Os03g53860 (rice) and At5g20950 (*Arabidopsis*), are drastically upregulated (>10-fold) from 3 HAI to 12 HAI (**Figure 4C and 4D**). Remarkably, rice ortholog (Os03g53860), associated with candidate gene GRMZM2G172043, is found in the same co-expression module as gene *RAmy1A*, the ortholog of our first discovered candidate gene, during early germination and imbibition in rice (Co-expression module E-MEXP-1766-Turquoise, $r^2 > 0.75$) (HOWELL *et al.* 2009). While these co-expression patterns are not to be unexpected given that both genes have been shown to be regulated by common hormone control (GA-activated), it is tempting to speculate if these cell-wall remodeling/carbohydrate metabolism genes exhibit significant epistatic variance in future scans. However, a previous comparison between the dry

seed transcriptomes of near isogenic lines (NILs) representing 'Delay of Germination' (*DOG)* quantitative trait loci (QTLs) of *Arabidopsis* that differ in after-ripening and/or dormancy suggests that natural variation for these traits are mainly controlled by additive genetic and molecular pathways, rather than epistatic interactions (BENTSINK *et al.* 2010).

Also noteworthy, maize protein levels of GRMZM2G172043 at 2 DAI (**Figure 4B**) are not nearly as upregulated as the transcriptional levels of the rice and *Arabidopsis* orthologs (**Figure 4C and 4D**), indicating that this gene may be 1) subject to post-transcriptional regulation 2) despite orthology, these genes are regulated differently across species or 3) the same phenomena is in fact occurring between species, but the lack of time-point resolution of maize seed tissues from 0-48 HAI prohibits the detection of transient gene expression patterns of this gene during early germination. There are several hypotheses from these data as to how β-glucanase activity may influence imbibition rate *K*. Differences in enzymatic activity rates may lead to differential cleavage of β-glucoside linkages and release of these molecules from the cell wall, leading to loosening, causing the water influx which drives cell expansion and generates cellular turgor pressure and subsequent embryo growth (SCHOPFER 2006). Indeed, a class I β-1,3-glucanase was found to specifically expressed in the micropylar endosperm of tobacco seeds prior to radicle emergence, and was suggested to be involved in endosperm weakening via cell-wall hydrolysis (LEUBNER-METZGER *et al.* 1995). Additionally, semi-permeability of mature cereal caryopses has been shown to be correlated with callose-layer thickness, a β-1,3-glucan substrate, on the inside of the seed coat (YIM AND BRADFORD 1998). Future experiments examining genotypic variation for physiology and anatomical structures, such as caryopsis thickness, metabolite deposition within various seed organs, air pocket spaces in the endosperm,

etc., along with reverse-genetics and mutant screens will be needed to confirm the fundamental seed characteristics which most underlie this trait.

Pathway-based enrichment analysis of biological processes associated with imbibition rate in the WiDiv association panel

SNP-based pathway enrichment analyses from GWAS results have been shown to be a promising approach to understand the genetic architecture and the over-arching molecular basis of quantitative traits in both animals and plants (WANG *et al.* 2010; CHEN *et al.* 2012; PEÑAGARICANO *et al.* 2012), This approach examines whether a group of related genes in the same functional pathway are jointly associated with a trait of interest, as GWAS alone often lacks the power to uncover the relatively small effect sizes conferred by multiple genetic variants (WANG *et al.* 2010). In our study, it was of particular interest to ascertain which biological pathways and/or specific tissues dictate imbibition rate *K*. While global pathways underlying early germination are comparatively more understood, especially in *Arabidopsis* relative to maize (NORTH *et al.* 2010; WEITBRECHT *et al.* 2011; DEKKERS *et al.* 2013), individual candidate genes associated with imbibition rate QTLs in other species never been identified via mapping or cloning, thus this trait currently remains a biological "black-box." Hence, we used pathway enrichment analysis on trait imbibition rate *K* via PlantGSEA (YI *et al.* 2013), interrogating three primary databases of gene lists [GO, KEGG, and classical gene families (GFam)] using an input gene list containing all genes with at least one significant SNP (based on a nominal p-value $\leq$ 0.005 from the previous GWAS of imbibition rate *K* (n = 1,147 genes).

In total, we identified 58 enriched biological terms from genes associated with imbibition rate *K*, with 36 GO terms (2 biological processes (BPs), 28 molecular functions (MFs), and 6 cellular compartments (CCs)), seven GFam terms, and 15 PlantCyc terms (**Table 3**). Remarkably, the majority of these biological terms have a previously implicated involvement in either seed development and/or early germination and share common biological pathways (WEITBRECHT *et al.* 2011). For instance, some of the most enriched terms include carbohydrate metabolic processes (GO: 000595, $q = 3.54x10^{-12}$), catalytic activity (GO:0003824, $q = 6.26x10^{-18}$), hydrolase activity (GO: 0016787, $q = 3.61x10^{-04}$), lactose degradation (PlantCyc, $q = 0.00158$), and sucrose degradation (PlantCyc, $q = 0.0415$), all of which are biological terms related to cell-wall degradation/remodeling or carbohydrate metabolism, which overlap terms with our most significant GWAS candidate genes, *ZmAMY1A* (α-amylase) and GRMZM2G172043 (periplasmic β-glucosidase), both of which are hydrolase enzymes (**Table 2 and Table 3**).

Other biological processes which have been shown to be universally upregulated during early imbibition (after 3 HAI) across species of seeds in both the endosperm and radicle of the embryonic axis include protein degradation/synthesis, ribosome assembly, translation associated processes, and nucleotide metabolism (DEKKERS *et al.* 2013; WALLEY *et al.* 2013). Terms over-represented in our pathway enrichment analysis related to these processes include protein binding (GO:0005488), ribosome (CC, GO:0005840), structural constituent of ribosome (GO:0003735), ubiquitin-protein ligase activity (GO:0004842), serine-type carboxypeptidase activity (GO:0004185), aspartic-type endopeptidase activity (GO:0004190), E3 F-box, E3 U-box, and E3 ring finger GFams associated with the ubiquitin-proteasome complex, and adenosine nucleotide degradation (PlantCyc) (**Table 3**). Previous experiments in *Arabidopsis* have

demonstrated that germination is a translation-dependent process, as seeds can still complete the germination process in the presence of a transcriptional inhibitor but cannot if translation is halted (RAJJOU *et al.* 2004). All components of the transcriptional machinery (including extant mRNAs) are stored in dry seeds and are quickly activated upon imbibition, as has been demonstrated by the fact that the addition of the translational inhibitor cycloheximide does not alter early transcript upregulation in Arabidopsis seeds during early germination, but it does disrupt early down-regulation of transcripts (KIMURA AND NAMBARA 2010). Translation thus seems to be essential to activate mRNA degradation mechanisms (WEITBRECHT *et al.* 2011). Interestingly, ribosomal proteins are among the first to be transcribed in both Arabidopsis and maize (BELTRÁN-PEÑA *et al.* 1995; TATEMATSU *et al.* 2008). Transcripts displaying strong changes in abundance from 1-3 HAI in rice are enriched for 3' untranslated regions (UTRs) with motifs associated with RNA stability, indicating that many changes are due to degradation of stored mRNAs (NAKABAYASHI *et al.* 2005; HOWELL *et al.* 2009). This is consistent with our pathway-analysis enrichment of adenosine nucleotide degradation (PlantCyc, q = 0.0396), including genes which degrade polyA tales of these UTRs. Drastic increases in amino acid content and decreases in proteins have been observed during the first six hours of imbibition in maize kernels (MEI AND SONG 2008), and it is tempting to speculate if increasing amino acid content and degradation of proteins could decrease osmotic potential and drive cell extension, and hence influence imbibition rate *K*.

Additionally, we observed an enrichment of biological terms associated with membranes and DNA repair (membrane and integral to membrane, GO_CC: 0016020/1, $q = 6.96 \times 10^{-06}$; helicase activity, GO_MF:0004386, $q = 9.85 \times 10^{-04}$; endonuclease activity, GO_MF: 0004519, $q = 5.77 \times 10^{-03}$) (**Table 3).** Drying of the seed and rehydration via imbibition imposes

considerable stresses upon the component cells resulting in temporary structural perturbations to membranes, which lead to an immediate and massive leakage of cellular solutes and metabolites, indicating damage to membranes and cellular compartments caused by fast and/or non-homogeneous rehydration (POWELL AND MATTHEWS 1978). This is symptomatic of a transition of the membrane phospholipid components from a "gel phase" achieved during maturing drying to the normal, hydrated liquid-crystalline state (CROWE AND CROWE 1992). In order to cope with the damage imposed during dehydration, storage, and most significantly, rehydration, seeds activate repair of membranes, proteins, and damaged genomic DNA, which results from the cumulative effects of temperature, moisture, oxygen, and ROS levels (BRAY AND WEST 2005). Insertional knock-out mutants of two DNA ligases, AtLIG4 and AtLIG6 consequently showed a delay in germination, confirming the necessity of repairing damage to DNA after traumatic imbibitional stress (WATERWORTH *et al.* 2010).

Furthermore, published transcriptomic and proteomic data sets indicates that different seed compartments, such as the endosperm, embryo, and caryopsis, accumulate unique transcripts and corresponding proteins during seed development (LE *et al.* 2010; WEITBRECHT *et al.* 2011; WALLEY *et al.* 2013). For example, proteins related to transcription factors and regulation of transcription were localized to the developing maize embryo, and kinases were abundant in the pericarp and aleurone (WALLEY *et al.* 2013). Pathway-enrichment analysis indicated a corresponding enrichment of genes related to protein kinase activity (GO_MF: 0004672, q =2.28 x 10^{-12}), regulation of transcription (GO_MF: 0030528, q =1.87 x 10^{-06}), transcription factor binding (GO_MF: 0003700, q = 3.61 x 10^{-04}), kinases (GFam, q =1.81 x 10^{-15}) , bHLH transcription factors (q =1.72x10^{-03}), and homeobox (HB) transcription factors (q = 0.0134) (**Table 3**). These two transcription factor families were also found to be enriched at the

proteomic level in a tissue-specific manner in the endosperm (bHLH) and embryo (HB), respectively (WALLEY *et al.* 2013). Importantly, Walley *et al.* 2013 characterized both the proteome and phosphoproteome of developing maize seeds and found that 1) correspondence between transcript and protein abundance in seeds is low (r = 0.41) for both endosperm and embryo tissues, and 2) protein abundance is independent of protein phosphorylation levels (WALLEY *et al.* 2013). These results collectively reinforce the necessity of interrogating gene function in seeds using a biological systems-based "multi-omics" approach due to their dynamic landscape of regulation.

Evaluating tissue-biased accumulation of non-modified protein and phosphorylation levels of top GWAS gene hits associated with imbibition rate

In order to better discern the seed tissues most responsible for controlling imbibition rate and to understand the regulation of these genes, we examined tissue-specific accumulation levels of both non-modified and phosphorylated proteins associated with our top 50 GWAS gene hits for *K* using the maize seed proteomic atlas (WALLEY *et al.* 2013). This seed proteomic atlas captures seed tissues such as the embryo, endosperm, endosperm cap, pericarp and aleurone during seed development, and also includes a germinating embryo sample 2 DAI. A total of 33 of these 50 (72%) GWAS gene hits express non-modified proteins in seed tissues (**Figure 5A**); this is significantly higher than expected by chance, considering that of the 23,899 genes containing RNA-seq derived SNPs interrogated by WiDiv GWAS, only 11,577 (48%) are expressed at the protein level. This observation supports that our GWAS results are fairly robust and are not merely a result of statistical noise. Non-modified proteins of these genes are mostly

expressed in a global manner across all developing seed organs (embryo, endosperm, endosperm cap, pericarp and aleurone, and germinating embryo), potentially indicating that all seed tissues contribute to imbibition rate in some manner (**Figure 5A,** panel I). These results are not unexpected, as only 1,203 of 12,453 surveyed non-modified in the maize seed proteome exhibit tissue-specific expression (9.6%) (WALLEY *et al.* 2013). After examining in which tissue these proteins exhibit maximum expression, the majority of genes do not exhibit peak protein expression in the germinating embryo, indicating that gene expression and translation during seed development are most likely influencing this trait as opposed to *de novo* transcription/translation induced upon imbibition. Additionally, a slight majority of gene candidates exhibit peak expression in the 27 DAP pericarp and aleurone tissue (7 of 33), with the next highest found in the endosperm crown (6 of 33), although peak gene expression is distributed relatively equal amongst seed tissues (**Figure 5B**). Given that the endosperm crown expresses the fewest number of proteins in the maize seed proteomic atlas, this may be a crucial tissue governing seed imbibition given that the endosperm cap is the region of the seed where both the radicle expands and pushes with osmotic force, and is subsequently weakened prior to radicle emergence and germination (MCDONALD 1994; WEITBRECHT *et al.* 2011). Walley *et al.* 2013 also found that contradictory to the non-modified proteome, phosphorylation levels of seed proteins are highly tissue dependent. We examined expression levels of phosphorylated polypeptides of our top 50 GWAS hits for *K* and found that 9 of these 36 (25%) proteins exist in a phosphorylated state (**Figure 5A**, panel II). Intriguingly, the phosphorylation status of most of these proteins exhibited tissue-specific accumulation, with the majority being highly phosphorylated in the germinating embryo (**Figure 5A**, panel II). This observation suggests that several imbibition gene candidates may seemingly lack expression in germinating embryos, but

may play key roles in the germination process via a dynamic, downstream signaling cascade of kinase phosphorylation and activation. Preliminary evidence suggests that control of imbibition rate in maize seeds is a complex interplay of several genes and seed tissues which are regulated at both the post-transcriptional and post-translational levels.

Evaluating relationships among seed imbibition and germination traits

In order to test the putative effect of imbibition rate *K* on agronomically-relevant germination characteristics, we selected 30 genotypes of the total 500 screened lines from the WiDiv population which displayed the substantial variation for imbibition rate *K*, sampling the ten fastest, slowest, and ten most intermediate genotypes from the distribution (**Appendix A5**). The phenotypic assay used to score two germination traits, total germination percentage after 72 H (%TG) and time to radicle emergence (TRE, given in hours) mirrored the agar plate scanner assay used to extract imbibition traits, as described previously (described further in the Materials and Methods section). Average genotypic imbibition rate *K* values of the 30 genotypes ranged from 35.4 (genotype CI90C) to 3.62 (genotype Il14H), with *K* group mean ± standard deviation values as follows: 31.6 ± 3.3 for the "fast" group, 19.5 ± 1.76 for the "medium" group, and 9.0 ± 3.4 for the "slow" group. Pairwise t-tests based on phenotypic values from both replications TRE and %TG were calculated between *K* groups to determine if significant differences existed. Remarkably, significant differences in germination performance were observed between all *K* groups for both TRE and %TG (**Figure 6**). The mean TRE for the "fast" *K* group was 43.7 hours, which was significantly different from the mean TRE values for both the slow (μ = 55.8, 95% confidence interval (CI): [-15.9, -8.2 h], $p = 2.19 \times 10^{-7}$) and medium (μ = 48.04, 95% CI: [-

7.6, -0.81 h], p = 0.016) (**Figure 6A**). The medium and slow *K* groups also significantly differed for TRE (95% CI: [-11.2, -4.4], $p = 4.32 \times 10^{-5}$). Pearson correlation analysis revealed that imbibition rate *K* and TRE (hours) were strongly negatively correlated in a linear fashion (r = -0.70, $p = 1.60 \times 10^{-5}$) amongst these extreme genotypes (**Figure 6B**). The linear equation TRE = -0.51(*K*) + 59.4 describing the model indicates that for every increase in one unit *K*, the time to radicle emergence is accelerated by about half an hour. Based on the differences in *K* values of our fastest (CI90C, *K*=35.4) and slowest genotypes (Il14H *K*=3.62), genotypic variation for TRE may vary by as much as 16.2 hours. While these differences between groups may seem minute, they may be more important in an applicable field setting where 1) water potentials are not 100% saturated, 2) soil temperatures are lower than the 20ºC room temperature used for the lab assay, and hence surrounding water potentials are decreased, and 3) the differences between *K* groups reported here may actually be an underestimate, given that the duration of the assay (72 hours) was not long enough for some of the slowest imbibing genotypes to complete the germination process.

Pairwise t-tests between *K* groups also revealed drastic differences between groups in respect to total percentage of germinated kernels after 72 hours (%TG). The mean %TG for the "fast" K group was 92.7%, which was significantly different from the mean TRE values for both the slow (μ = 61.6% 95% confidence interval (CI): [19%, 42%], $p = 4.6 \times 10^{-6}$) and medium (μ = 83.3%, 95% CI: [1.3%, 17.6%], p = 0.02) (**Figure 6C**). The medium and slow *K* groups also varied for %TG (95% CI: [10%, 33.3%, p = 0.0006). Pearson correlation analysis revealed that imbibition rate *K* and %TG were significantly positively correlated (r = 0.70, p = 1.66×10^{-5}) amongst these diverse genotypes (**Figure 6D**). Correlations for all measured seed and germination traits of these 30 extreme genotypes are shown in **Appendix A12**. The linear

equation %TG = 0.013(*K*) + 54.1 describing the model indicates that for every increase in one unit *K*, %TG is increased by 1.3%. In contrast to TRE, the estimates in differences between groups for %TG may be over-estimated, as the duration of the assay was not long enough for the slowest imbibing genotypes to complete germination.

The uptake of water by the seed during germination follows a classic triphasic model, commencing with rapid initial water uptake (Phase I, i.e. imbibition), followed by a plateau "lag" phase (Phase II). It is the length of the "lag period" that slows germination rates via extended dormancy, low temperature, water deficit, or ABA, while factors which accelerate germination do so by shortening the lag phase (BRADFORD 1990). However, the length and speed of transition to Phase II of the germination process is presumably related to the generation of additional embryo turgor, embryo cell-wall loosening, or weakening of tissues enclosing the embryo (WELBAUM *et al.* 1998). Previous research in lettuce has found that germination rate increased linearly with embryo turgor (Ψ_{embryo}) (BRADFORD 1990). These researchers demonstrated that the endosperm presented little physical resistance to radicle growth at the time of radicle emergence, but its presence markedly delayed germination. Importantly, it was found that the length of the lag period after imbibition before radicle emergence was primarily related to the time required for weakening of the endosperm and not the generation of additional turgor in the embryo (BRADFORD 1990). Cracks and breaks between endosperm cells have been observed just prior to visible radicle growth, with the expanding radicle tip apparently forcing the endosperm tissue apart along these cracks. This indicates that Ψ_{embryo} influences the rate at which endosperm weakening occurs (BRADFORD 1990). Additionally, a previously generated maize seed imbibition model revealed that the proportion of the imbibed seed surface was strongly correlated with the seed imbibition rate and consequently with the germination rate

(BRUCKLER 1983; GARDARIN *et al.* 2011). Hence, drastic variation in imbibition rates may hasten the remobilization of seed reserves and subsequent germination rates amongst diverse cultivars.

In general, across both traits, the medium group was more similar in performance to the fast group, suggesting a threshold of ideal imbibition rate *K* which is necessary for optimal germination rate and percentage. While the observed significant relationship between *K* and important germination characteristics is extremely interesting, further tests with additional genotypes will be needed to confirm this association. Nevertheless, these results indicate that genotypic diversity for imbibition rate *K* may be one physiological mechanism dictating the variation in germination traits amongst cultivars. Given the prior report that increasing embryo turgor is linearly correlated to germination rate (BRADFORD 1990), it is tempting to speculate if *K* is capturing this phenomenon as well, or if higher water concentrations in the seed at earlier time points simply activates cell-wall loosening enzymes and remobilization of stored energy reserves at a faster rate, leading to a quicker transition towards Phase II of the germination process (WELBAUM *et al.* 1998). Additionally, our results indicate that our phenotyping platform for imbibition rate *K* could serve as an applicable, high-throughput and accurate method to quickly screen genotypes for optimal germination and emergence rates in maize.

3.4 CONCLUSIONS

The goal of our study was to characterize the genetic architecture and molecular networks dictating seed imbibition and its relationship to germination in maize. We demonstrated that significant phenotypic and genetic variation and subsequent high repeatabilities exist for seed imbibition traits. In this study, a total of a total of 104 trait-associative SNPs (encompassing 67 genes) exceeding the genome-wide suggestive threshold were identified for three seed and imbibition traits in maize. Three genetic regions associated with imbibition rate constant *K* were identified via GWA. Using a systems biology multi-omics approach which integrated genomic, transcriptomic, and proteomic data sets, we were able to resolve these imbibition rate associated genomic regions to identify two putative candidate genes involved in cell-wall breakdown and hydrolysis: *ZmAMY1A* (GRMZM2G103055), which encodes for a putative α-amylase protein, and GRMZM2G172043, a putative β-1,3-glucanase protein. Quantitative trait pathway analysis of genes identified from the GWAS results showed enrichment of biological processes related to carbohydrate metabolism, catalytic and hydrolase activity, protein degradation, translation, membrane structure, nucleotide metabolism, and kinases involved in signaling cascades. These terms overlapped with both our identified candidate genes and global processes known to be differentially expressed during imbibition and early germination in other plant species. Clustering of seed proteomic levels using top GWAS gene hits revealed genes associated with imbibition function globally in all seed tissues but exhibit tissue-specific phosphorylation. Using a subset of extreme genotypes, we demonstrated a highly significant correlation between *K* and radicle emergence time and germination percentage, indicating that further characterization of the genetics and physiology underlying seed imbibition may be useful to enhance the early vigor and performance of maize.

LITERATURE CITED

Akazawa, T., J. Yamaguchi and M. Hayashi, 1991 Rice α-amylase and gibberellin action—a personal view, pp. 114-124 in *Gibberellins*. Springer.

Altelaar, A. M., J. Munoz and A. J. Heck, 2013 Next-generation proteomics: towards an integrative view of proteome dynamics. Nature Reviews Genetics 14**:** 35-48.

Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler *et al.*, 2000 Gene Ontology: tool for the unification of biology. Nature genetics 25**:** 25-29.

Avraham, S., C.-W. Tung, K. Ilic, P. Jaiswal, E. A. Kellogg *et al.*, 2008 The Plant Ontology Database: a community resource for plant structure and developmental stages controlled vocabulary and annotations. Nucleic acids research 36**:** D449-D454.

Bates, D., B. Bolker and S. Walker, 2015 lme4: Linear mixed-effects models using S4 classes., pp. R Core Team.

Bathgate, G., G. Palmer and G. Wilson, 1974 The action of endo-β-1,3-glucanases on barley and malt β-glucans. Journal of the Institute of Brewing 80**:** 278-285.

Beltrán-Peña, E., A. Ortíz-López and E. S. de Jiménez, 1995 Synthesis of ribosomal proteins from stored mRNAs early in seed germination. Plant molecular biology 28**:** 327-336.

Benjamini, Y., and Y. Hochberg, 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological)**:** 289-300.

Bentsink, L., J. Hanson, C. J. Hanhart, H. Blankestijn-de Vries, C. Coltrane *et al.*, 2010 Natural variation for seed dormancy in Arabidopsis is regulated by additive genetic and molecular pathways. Proceedings of the National Academy of Sciences 107**:** 4264-4269.

Bewley, J. D., 1997 Seed germination and dormancy. The Plant Cell 9**:** 1055.

Bewley, J. D., and M. Black, 1994 *Seeds*. Springer.

Bradford, K. J., 1986 Manipulation of seed water relations via osmotic priming to improve germination under stress conditions. HortScience (USA).

Bradford, K. J., 1990 A water relations analysis of seed germination rates. Plant Physiology 94**:** 840-849.

Bray, C. M., and C. E. West, 2005 DNA repair mechanisms in plants: crucial sensors and effectors for the maintenance of genome integrity. New Phytologist 168**:** 511-528.

Bruckler, L., 1983 Rôle des propriétés physiques du lit de semences sur l'imbibition et la germination. II. Contrôle expérimental d'un modèle d'imbibition des semences et possibilités d'applications. Agronomie 3**:** 223-232.

Chen, C., G. DeClerck, F. Tian, W. Spooner, S. McCouch *et al.*, 2012 PICARA, an analytical pipeline providing probabilistic inference about a priori candidates genes underlying genome-wide association QTL in plants.

Crowe, J., and L. Crowe, 1992 Membrane integrity in anhydrobiotic organisms: toward a mechanism for stabilizing dry cells, pp. 87-103 in *Water and Life*. Springer.

Dekkers, B. J., S. Pearce, R. van Bolderen-Veldkamp, A. Marshall, P. Widera *et al.*, 2013 Transcriptional dynamics of two seed compartments with opposing roles in Arabidopsis seed germination. Plant physiology 163**:** 205-215.

Dias, P. M. B., S. Brunel-Muguet, C. Dürr, T. Huguet, D. Demilly *et al.*, 2011 QTL analysis of seed germination and pre-emergence growth at extreme temperatures in Medicago truncatula. Theoretical and Applied Genetics 122**:** 429-444.

Dure, L. S., 1960 Site of origin and extent of activity of amylases in maize germination. Plant physiology 35**:** 925.

Falconer, D. S., and T. F. C. Mackay, 1996 *Introduction to quantitative genetics*. Prentice Hall, New York, New York, USA, Pearson Education Limited.

Galland, M., R. Huguet, E. Arc, G. Cueff, D. Job *et al.*, 2014 Dynamic proteomics emphasizes the importance of selective mRNA translation and protein turnover during Arabidopsis seed germination. Molecular & Cellular Proteomics 13**:** 252-268.

Gao, X., 2011 Multiple testing corrections for imputed SNPs. Genetic epidemiology 35**:** 154-158.

Gao, X., L. C. Becker, D. M. Becker, J. D. Starmer and M. A. Province, 2010 Avoiding the high Bonferroni penalty in genome-wide association studies. Genetic epidemiology 34**:** 100-105.

Gao, X., J. Starmer and E. R. Martin, 2008 A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. Genetic epidemiology 32**:** 361.

Gardarin, A., C. Dürr and N. Colbach, 2011 Prediction of germination rates of weed species: relationships between germination speed parameters and species traits. Ecological Modelling 222**:** 626-636.

Gubler, F., R. Kalla, J. K. Roberts and J. V. Jacobsen, 1995 Gibberellin-regulated expression of a myb gene in barley aleurone cells: evidence for Myb transactivation of a high-pI alpha-amylase gene promoter. The Plant Cell 7**:** 1879-1891.

Gutterman, Y., 2000 Maternal effects on seeds during development. Seed: the ecology of regeneration in plant communities**:** 59-84.

Hansey, C. N., J. M. Johnson, R. S. Sekhon, S. M. Kaeppler and N. d. Leon, 2011 Genetic diversity of a maize association population with restricted phenology. Crop Science 51**:** 704-715.

Hawkins, R. D., G. C. Hon and B. Ren, 2010 Next-generation genomics: an integrative approach. Nature Reviews Genetics 11**:** 476-486.

Hirsch, C. N., J. M. Foerster, J. M. Johnson, R. S. Sekhon, G. Muttoni *et al.*, 2014 Insights into the maize pan-genome and pan-transcriptome. The Plant Cell Online 26**:** 121-135.

Høj, P. B., and G. B. Fincher, 1995 Molecular evolution of plant β-glucan endohydrolases. The Plant Journal 7**:** 367-379.

Howell, K. A., R. Narsai, A. Carroll, A. Ivanova, M. Lohse *et al.*, 2009 Mapping metabolic and transcript temporal switches during germination in rice highlights specific transcription factors and the role of RNA instability in the germination process. Plant physiology 149**:** 961-980.

Howie, B. N., P. Donnelly and J. Marchini, 2009 A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet 5**:** e1000529.

Hrmova, M., and G. B. Fincher, 2001 Structure-function relationships of β-D-glucan endo-and exohydrolases from higher plants. Plant molecular biology 47**:** 73-91.

Ingvarsson, P. K., and N. R. Street, 2011 Association genetics of complex traits in plants. New Phytologist 189**:** 909-922.

Jones, R. L., and J. V. Jacobsen, 1991 Regulation of synthesis and transport of secreted proteins in cereal aleurone. International review of cytology (USA).

Kanehisa, M., and S. Goto, 2000 KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research 28**:** 27-30.

Kanehisa, M., S. Goto, S. Kawashima, Y. Okuno and M. Hattori, 2004 The KEGG resource for deciphering the genome. Nucleic acids research 32**:** D277-D280.

Kaneko, M., H. Itoh, M. Ueguchi-Tanaka, M. Ashikari and M. Matsuoka, 2002 The α-amylase induction in endosperm during rice seed germination is caused by gibberellin synthesized in epithelium. Plant Physiology 128**:** 1264-1270.

Karrer, E. E., J. M. Chandler, M. R. Foolad and R. L. Rodriguez, 1992 Correlation between α-amylase gene expression and seedling vigor in rice. Euphytica 66**:** 163-169.

Karrer, E. E., and R. L. Rodriguez, 1992 Metabolic regulation of rice α-amylase and sucrose synthase genes in planta. The Plant Journal 2**:** 517-523.

Kimura, M., and E. Nambara, 2010 Stored and neosynthesized mRNA in Arabidopsis seeds: effects of cycloheximide and controlled deterioration treatment on the resumption of transcription during imbibition. Plant molecular biology 73**:** 119-129.

Korte, A., and A. Farlow, 2013 The advantages and limitations of trait analysis with GWAS: a review. Plant methods 9**:** 29.

Laird, P. W., 2010 Principles and challenges of genome-wide DNA methylation analysis. Nature Reviews Genetics 11**:** 191-203.

Le, B. H., C. Cheng, A. Q. Bui, J. A. Wagmaister, K. F. Henry *et al.*, 2010 Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors. Proceedings of the National Academy of Sciences 107**:** 8063-8070.

Leopold, A. C., 1983 Volumetric Components of Seed Imbibition. Plant Physiology 73**:** 677-680.

Leubner-Metzger, G., 2003 Functions and regulation of β-1,3-glucanases during seed germination, dormancy release and after-ripening. Seed Science Research 13**:** 17-34.

Leubner-Metzger, G., C. Frundt, R. Vogeli-Lange and F. Meins Jr, 1995 Class I [beta]-1, 3-Glucanases in the Endosperm of Tobacco during Germination. Plant Physiology 109**:** 751-759.

Lipka, A. E., F. Tian, Q. Wang, J. Peiffer, M. Li *et al.*, 2012 GAPIT: genome association and prediction integrated tool. Bioinformatics 28**:** 2397-2399.

Liptay, A., and P. Schopfer, 1983 Effect of water stress, seed coat restraint, and abscisic acid upon different germination capabilities of two tomato lines at low temperature. Plant Physiology 73**:** 935-938.

Martin, B., O. Smith and M. O'neil, 1988 Relationships between laboratory germination tests and field emergence of maize inbreds. Crop Science 28**:** 801-805.

McDonald, M. B., 1994 Seed germination and seedling establishment. Physiology and determination of crop yield**:** 37-60.

McLaren, W., Pritchard, B., Rios, D., Chen, Y., Flicek, P., and Cunningham, F. (2010). Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. Bioinformatics 26: 2069-2070.

Mei, Y.-Q., and S.-Q. Song, 2008 Early morphological and physiological events occurring during germination of maize seeds. Agricultural Sciences in China 7**:** 950-957.

Metzker, M. L., 2010 Sequencing technologies—the next generation. Nature reviews genetics 11**:** 31-46.

Monaco, M. K., Stein, J., Sushma, N., Wei, S., Dharmawardhana, P., Sunita, K., *et al.* Gramene 2013: comparative plant genomics resources. 2014 Gramene 2013: comparative plant genomics resources. Nucleic acids research 42(D1): D1193-D1199.

Nakabayashi, K., M. Okamoto, T. Koshiba, Y. Kamiya and E. Nambara, 2005 Genome-wide profiling of stored mRNA in Arabidopsis thaliana seed germination: epigenetic and genetic regulation of transcription in seed. The plant journal 41**:** 697-709.

Noggle, G. R., and G. J. Fritz, 1983 *Introductory plant physiology*. Prentice-Hall Inc.

Nordborg, M., and D. Weigel, 2008 Next-generation genetics in plants. Nature 456**:** 720-723.

North, H., S. Baud, I. Debeaujon, C. Dubos, B. Dubreucq *et al.*, 2010 Arabidopsis seed secrets unravelled after a decade of genetic and omics-driven research. The Plant Journal 61**:** 971-981.

O'Neil, S., M. Kumagai, A. Majumdar, N. Hunang, T. Sutliff *et al.*, 1990 The alpha-amylase genes in Oryza sativa: characterization of cDNA clones and mRNA expression during seed development. Mol Gen Genet 221**:** 235-244.

Ozsolak, F., and P. M. Milos, 2011 RNA sequencing: advances, challenges and opportunities. Nature reviews genetics 12**:** 87-98.

Park, P. J., 2009 ChIP–seq: advantages and challenges of a maturing technology. Nature Reviews Genetics 10**:** 669-680.

Peñagaricano, F., K. A. Weigel, G. J. Rosa and H. Khatib, 2012 Inferring quantitative trait pathways associated with bull fertility from a genome-wide association study. Frontiers in genetics 3.

Perata, P., N. Geshi, J. Yamaguchi and T. Akazawa, 1993 Effect of anoxia on the induction of α-amylase in cereal seeds. Planta 191**:** 402-408.

Powell, A. A., and S. Matthews, 1978 The damaging effect of water on dry pea embryos during imbibition. Journal of Experimental Botany 29**:** 1215-1229.

Preston, J., K. Tatematsu, Y. Kanno, T. Hobo, M. Kimura *et al.*, 2009 Temporal expression patterns of hormone metabolism genes during imbibition of Arabidopsis thaliana seeds: a comparative study on dormant and non-dormant accessions. Plant and cell physiology 50**:** 1786-1800.

R Development Core Team, 2013 R: A language and environment for statistical computing., pp. R Foundation for Statistical Computing, Vienna, Austria.

Rajjou, L., M. Duval, K. Gallardo, J. Catusse, J. Bally *et al.*, 2012 Seed germination and vigor. Annual review of plant biology 63**:** 507-533.

Rajjou, L., K. Gallardo, I. Debeaujon, J. Vandekerckhove, C. Job *et al.*, 2004 The effect of α-amanitin on the Arabidopsis seed proteome highlights the distinct roles of stored and neosynthesized mRNAs during germination. Plant Physiology 134**:** 1598-1613.

Remington, D. L., J. M. Thornsberry, Y. Matsuoka, L. M. Wilson, S. R. Whitt *et al.*, 2001 Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proceedings of the National Academy of Sciences 98**:** 11479-11484.

Ritchie, M. D., E. R. Holzinger, R. Li, S. A. Pendergrass and D. Kim, 2015 Methods of integrating data to uncover genotype-phenotype interactions. Nature Reviews Genetics 16**:** 85-97.

Roach, D. A., and R. D. Wulff, 1987 Maternal effects in plants. Annual review of ecology and systematics**:** 209-235.

Robert, C., A. Noriega, Á. Tocino and E. Cervantes, 2008 Morphological analysis of seed shape in Arabidopsis thaliana reveals altered polarity in mutants of the ethylene signaling pathway. Journal of plant physiology 165**:** 911-919.

Sanwo, M. M., and D. A. DeMason, 1992 Characteristics of α-amylase during germination of two high-sugar sweet corn cultivars of *Zea mays L*. Plant physiology 99**:** 1184-1192.

Sanwo, M. M., and D. A. deMason, 1993 A Comparison of α-Amylase Isozyme Profiles in Selected Su and High-Sugar Sweet Corn (sh-2, su-1, su-1 se) Lines (*Zea mays L.*). International journal of plant sciences**:** 395-405.

Schnable, J. C., M. Freeling and E. Lyons, 2012 Genome-wide analysis of syntenic gene deletion in the grasses. Genome biology and evolution 4**:** 265-277.

Schopfer, P., 2006 Biomechanics of plant growth. American journal of botany 93**:** 1415-1425.

Schopfer, P., and C. Plachy, 1984 Control of Seed Germination by Abscisic Acid II. Effect on Embryo Water Uptake in Brassica napus L. Plant Physiology 76**:** 155-160.

Schwämmle, V., T. Verano-Braga and P. Roepstorff, 2015 Computational and statistical methods for high-throughput analysis of post-translational modifications of proteins. Journal of proteomics 129**:** 3-15.

Shaw, R. H., and W. Loomis, 1950 Bases for the prediction of corn yields. Plant Physiology 25**:** 225.

Shaykewich, C., 1973 Proposed method for measuring swelling pressure of seeds prior to germination. Journal of Experimental Botany 24**:** 1056-1061.

Shin, J.-H., S. Blay, B. McNeney and J. Graham, 2006 LDheatmap: an R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. Journal of Statistical Software 16**:** 1-10.

Shulaev, V., 2006 Metabolomics technology and bioinformatics. Briefings in Bioinformatics 7**:** 128-139.

Simmons, C. R., 1994 The physiology and molecular biology of plant 1, 3-β-D-glucanases and 1, 3; 1, 4-β-D-glucanases. Critical Reviews in Plant Sciences 13**:** 325-387.

Skriver, K., F. L. Olsen, J. C. Rogers and J. Mundy, 1991 cis-acting DNA elements responsive to gibberellin and its antagonist abscisic acid. Proceedings of the National Academy of Sciences 88**:** 7266-7270.

Stelpflug, S., R. S. Sekhon, B. Vaillancourt, C. N. Hirsch, C. R. Buell *et al.*, 2015 An expanded maize gene expression atlas based on RNA-sequenceing and its use to explore root development. The Plant Genome, doi 10.

Streiner, D. L., and G. R. Norman, 2011 Correction for multiple testing: is there a resolution? CHEST Journal 140**:** 16-18.

Sugimoto, N., G. Takeda, Y. Nagato and J. Yamaguchi, 1998 Temporal and spatial expression of the α-amylase gene during seed germination in rice and barley. Plant and cell physiology 39**:** 323-333.

Tatematsu, K., Y. Kamiya and E. Nambara, 2008 Co-regulation of ribosomal protein genes as an indicator of growth status: comparative transcriptome analysis on axillary shoots and seeds in Arabidopsis. Plant signaling & behavior 3**:** 450-452.

Tavazoie, S., J. D. Hughes, M. J. Campbell, R. J. Cho and G. M. Church, 1999 Systematic determination of genetic network architecture. Nature genetics 22**:** 281-285.

Tenaillon, M. I., M. C. Sawkins, A. D. Long, R. L. Gaut, J. F. Doebley *et al.*, 2001 Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays L.*). Proceedings of the National Academy of Sciences 98**:** 9161-9166.

Upton, G. J., 1992 Fisher's exact test. Journal of the Royal Statistical Society. Series A (Statistics in Society)**:** 395-402.

VanRaden, P., 2008 Efficient methods to compute genomic predictions. Journal of dairy science 91**:** 4414-4423.

Vatandoust, A., S. Ragaee, P. J. Wood, S. M. Tosh and K. Seetharaman, 2012 Detection, localization, and variability of endogenous β-glucanase in wheat kernels. Cereal Chemistry 89**:** 59-64.

Vertucci, C. W., 1989 The effects of low water contents on physiological activities of seeds. Physiologia Plantarum 77**:** 172-176.

Walley, J. W., Z. Shen, R. Sartor, K. J. Wu, J. Osborn *et al.*, 2013 Reconstruction of protein networks from an atlas of maize seed proteotypes. Proceedings of the National Academy of Sciences 110**:** E4808-E4817.

Wang, K., M. Li and H. Hakonarson, 2010 Analysing biological pathways in genome-wide association studies. Nature Reviews Genetics 11**:** 843-854.

Wang, Z., M. Gerstein and M. Snyder, 2009 RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics 10**:** 57-63.

Waters, A. J., I. Makarevitch, S. R. Eichten, R. A. Swanson-Wagner, C.-T. Yeh *et al.*, 2011 Parent-of-origin effects on gene expression and DNA methylation in the maize endosperm. The Plant Cell 23**:** 4221-4233.

Waterworth, W. M., G. Masnavi, R. M. Bhardwaj, Q. Jiang, C. M. Bray *et al.*, 2010 A plant DNA ligase is an important determinant of seed longevity. The Plant Journal 63**:** 848-860.

Weitbrecht, K., K. Müller and G. Leubner-Metzger, 2011 First off the mark: early seed germination. Journal of experimental botany 62**:** 3289-3309.

Welbaum, G., K. Bradford, K.-O. Yim, D. Booth and M. Oluoch, 1998 Biophysical, physiological and biochemical processes regulating seed germination. Seed Science Research 8**:** 161-172.

Woodstock, L., 1988 Seed imbibition: a critical period for successful germination. Journal of Seed Technology**:** 1-15.

Yang, W., Z. Guo, C. Huang, L. Duan, G. Chen *et al.*, 2014 Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. Nature communications 5.

Yi, X., Z. Du and Z. Su, 2013 PlantGSEA: a gene set enrichment analysis toolkit for plant community. Nucleic acids research 41**:** W98-W103.

Yim, K.-O., and K. J. Bradford, 1998 Callose deposition is responsible for apoplastic semipermeability of the endosperm envelope of muskmelon seeds. Plant Physiology 118**:** 83-90.

Young, T., J. Juvik and D. DeMason, 1997 Changes in carbohydrate composition and α-amylase expression during germination and seedling growth of starch-deficient endosperm mutants of maize. Plant Science 129**:** 175-189.

Yu, J., G. Pressoir, W. H. Briggs, I. V. Bi, M. Yamasaki *et al.*, 2006 A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature genetics 38**:** 203-208.

Yu, S.-M., Y.-C. Lee, S.-C. Fang, M.-T. Chan, S.-F. Hwa *et al.*, 1996 Sugars act as signal molecules and osmotica to regulate the expression of α-amylase genes and metabolic activities in germinating cereal grains. Plant molecular biology 30**:** 1277-1289.

Zhang, P., K. Dreher, A. Karthikeyan, A. Chi, A. Pujar *et al.*, 2010 Creation of a genome-wide metabolic pathway database for Populus trichocarpa using a new approach for reconstruction and curation of metabolic pathways for plants. Plant physiology 153**:** 1479-1491.

Zwar, J. A., and R. Hooley, 1986 Hormonal regulation of α-amylase gene transcription in wild oat (Avena fatua L.) aleurone protoplasts. Plant physiology 80**:** 459-463.

Table 1. Repeatability estimates, genetic variance components, and Spearman's rank correlations of maize seed and imbibition traits measured in the WiDiv association panel. The WiDiv association panel was evaluated with a RCBD experimental design with two replications blocked in time (n=500). Genotypic variance (σ^2_G) was significant for all phenotypic traits ($p < 0.001$). Spearman's rank correlations were calculated for each trait between replicates. *R*: repeatability, σ^2_e: residual variance, *K*: imbibition rate constant, %ISSA: percentage increase of seed surface area after 24 hours imbibition. *** significant at $P < 0.0001$

Trait	σ^2_G	σ^2_e	*R* (%)	Spearman's rank (ρ)
K	18.99	5.53	87.3	0.75 ***
%ISSA	0.0009094	0.000404	81.8	0.64 ***
Seed size	3.97E+08	5.42E+07	93.6	0.86 ***

Table 2. Trait-associated SNPs from GWAS results for maize seed and imbibition traits measured in the WiDiv association panel (*n*=500). TAS, trait-associated SNP; Chr: chromosome; maf: minor allele frequency; % Var. exp.: variance explained by the TAS (in %). Positions are based on AGPv2 (in base pairs). *K*: imbibition rate; %ISSA: percentage increase in seed surface area after 24H.

Trait	TAS	Chr.	Position	*p*-value	maf	% Var. exp.	Gene Model	Annotation
K	rna5_209899708	5	209899708	8.61E-07	0.396	4.3%	GRMZM2G103055	Alpha-amylase *AMY1*
K	rna5_209899702	5	209899702	1.37E-06	0.393	4.1%	GRMZM2G103055	Alpha-amylase *AMY1*
K	rna5_7241582	5	7241582	3.35E-06	0.088	3.8%	GRMZM2G172043	Periplasmic beta-glucosidase
K	rna3_13484774	3	13484774	4.17E-06	0.04	3.7%	GRMZM5G858165	NET4B, KIP-1like
%ISSA	rna5_12552026	5	12552026	3.65E-12	0.012	7.9%	GRMZM2G327260	Unknown function
%ISSA	rna5_1549707	5	1549707	9.00E-10	0.029	6.1%	GRMZM2G134270	FAD/NAD(P)-binding oxidoreductase
%ISSA	rna5_116868745	5	116868745	1.10E-09	0.022	6.0%	GRMZM2G171604	*prc4*, proteosome component 4
%ISSA	rna5_144812971	5	144812971	1.90E-09	0.022	5.8%	GRMZM2G069542	*pep2*, phosphoenolpyruvate carboxylase 2
%ISSA	rna5_144812976	5	144812976	1.90E-09	0.022	5.8%	GRMZM2G069542	*pep2*, phosphoenolpyruvate carboxylase 2
%ISSA	rna2_1977417	2	1977417	3.35E-09	0.031	5.6%	GRMZM2G068059	DRE2, Cytokine-induced anti-apoptosis inhibitor
%ISSA	rna8_116310218	8	116310218	1.05E-08	0.044	5.3%	GRMZM2G055489	*spp1*, sucrose phosphatase 1
%ISSA	rna9_127427853	9	127427853	1.71E-08	0.046	5.1%	GRMZM2G105523	Pyrrolidone-carboxylate peptidase
%ISSA	rna4_156022733	4	156022733	3.03E-08	0.02	4.9%	GRMZM2G028438	*ZmGRAS76*
%ISSA	rna4_156022735	4	156022735	3.03E-08	0.02	4.9%	GRMZM2G028438	*ZmGRAS76*
%ISSA	rna4_50157636	4	50157636	3.35E-08	0.046	4.9%	GRMZM2G177052	auxin-regulated protein DUF966
%ISSA	rna9_146206275	9	146206275	3.55E-08	0.028	4.9%	GRMZM5G824920	Glucan endo-1,3-beta-glucosidase 3
%ISSA	rna9_138707816	9	138707816	1.05E-07	0.053	4.5%	GRMZM2G011071	*ABCI16*, ATP-Binding Cassette I16
%ISSA	rna1_147014236	1	147014236	1.20E-07	0.048	4.5%	GRMZM2G147724	Inositol phosphorylceramide synthase
%ISSA	rna4_156481266	4	156481266	1.70E-07	0.03	4.4%	GRMZM2G084195	Histone H4
%ISSA	rna4_156481269	4	156481269	1.70E-07	0.03	4.4%	GRMZM2G084195	Histone H4
%ISSA	rna6_2721850	6	2721850	1.75E-07	0.04	4.4%	GRMZM2G367941	SNARE-like superfamily protein
%ISSA	rna1_188032064	1	188032064	1.94E-07	0.024	4.3%	GRMZM5G879527	*ZmbHLH112*
%ISSA	rna3_5596727	3	5596727	2.23E-07	0.012	4.3%	GRMZM2G143165	Alkaline-phosphatase-like protein

Table 3. Significantly over-represented biological terms for genes associated with seed imbibition rate *K* ($P < 0.005$, n=1147 genes). Gene sets from three databases were interrogated: Gene Ontology (GO), curated gene family (GFam) lists, and PlantCyc. Only the databases for which we found significant results ($q \leq 0.05$) are shown. BP: biological processes; MF: molecular function; CC: cellular compartment; FDR: False Discovery Rate.

GO Term ID	Biological Description	# of background genes	# of input list genes	*q*-value
	GO Biological Process			
GO:0008152	Metabolic process	1021	71	1.31E-07
GO:0005975	Carbohydrate metabolic process	452	35	3.54E-12
	GO Molecular Function			
GO:0003824	Catalytic activity	1008	67	6.26E-18
GO:0004672	Protein kinase activity	975	56	2.28E-12
GO:0004674	Protein serine/threonine kinase activity	996	56	3.54E-12
GO:0005515	Protein binding	1395	61	4.87E-09
GO:0016491	Oxioreductase activity	696	38	1.09E-07
GO:0005488	Binding	491	31	1.54E-07
GO:0004553	Hydrolase activity (O-glycosyl compounds)	227	20	8.14E-07
GO:0030528	Transcriptional regulator activity	293	22	1.87E-06
GO:0020037	Heme binding	380	24	8.10E-06
GO:0005215	Transporter activity	332	22	1.22E-05
GO:0000166	Nucleotide binding	487	27	1.24E-05
GO:0003735	Structural constituent of ribosome	491	27	1.34E-05
GO:0004842	Ubiquitin-protein ligase activity	77	11	2.02E-05
GO:0046872	Metal ion binding	195	16	3.83E-05
GO:0003676	Nucleic acid binding	736	33	3.97E-05
GO:0030170	Pyridoxal phosphate binding	94	11	9.58E-05
GO:0003700	Sequence-specific DNA binding transcription factors	529	25	3.61E-04
GO:0016787	Hydrolase activity	269	17	3.61E-04
GO:0017111	Nucleoside triphosphate activity	390	20	9.79E-04
GO:0004386	Helicase activity	125	11	9.85E-04
GO:0016758	Transferase activity	139	11	2.32E-03
GO:0009055	Electron carrier activity	411	19	5.11E-03
GO:0004185	Serine-type carboxypeptidase activity	39	6	5.28E-03
GO:0004519	Endonuclease activity	40	6	5.77E-03
GO:0016887	ATPase activity	167	11	9.10E-03

GO Molecular Function (cont.)				
GO:0004190	Aspartic-type endopeptidase activity	73	7	0.0174
GO:0005509	Calcium ion binding	223	12	0.0243
GO:0004601	Peroxidase activity	132	9	0.0251
GO Cellular Compartment				
GO:0005634	Nucleus	1172	61	6.96E-06
GO:0016020	Membrane	1364	68	6.96E-06
GO:0016021	Integral to membrane	967	51	3.94E-05
GO:0000151	Ubiquitin ligase complex	75	11	3.55E-04
GO:0005622	Intracellular	1141	51	1.83E-03
GO:0005840	Ribosome	483	26	9.13E-03
Gene Families				
	Kinase	841	48	1.81E-15
	E3 Ubox domain-containing family	54	8	8.03E-05
	bHLH transcription factors	94	8	1.72E-03
	Auxin/IAA gene family	26	5	1.72E-03
	Homeobox TFs	73	6	0.0134
	E3 Fbox family	162	8	0.0324
	E3 ring finger family	372	12	0.049
PlantCyc				
	Glycolysis III	138	13	1.09E-06
	Glycolysis I	137	13	1.09E-06
	Glycolysis IV (Plant cytosol)	128	13	1.09E-06
	Gluconeogenesis	98	10	2.17E-05
	Adenosine nucleotide degradation I	28	6	1.46E-04
	Suberin biosynthesis	26	5	1.49E-03
	Lactose degradation III	12	4	1.58E-03
	Phenylpropanoid biosynthesis	39	5	2.01E-03
	Brassinosteroid biosynthesis	82	6	0.0163
	Triacylglycerol degradation	81	6	0.0163
	Flavanoid and anthocyanin biosynthesis	56	5	0.022
	Pentose phosphate pathway (non-oxidative branch)	14	3	0.0329
	ABA biosynthesis	16	3	0.0396
	Adenosine nucleotides *de novo* biosynthesis	106	6	0.0396
	Sucrose degradation III	71	5	0.0415

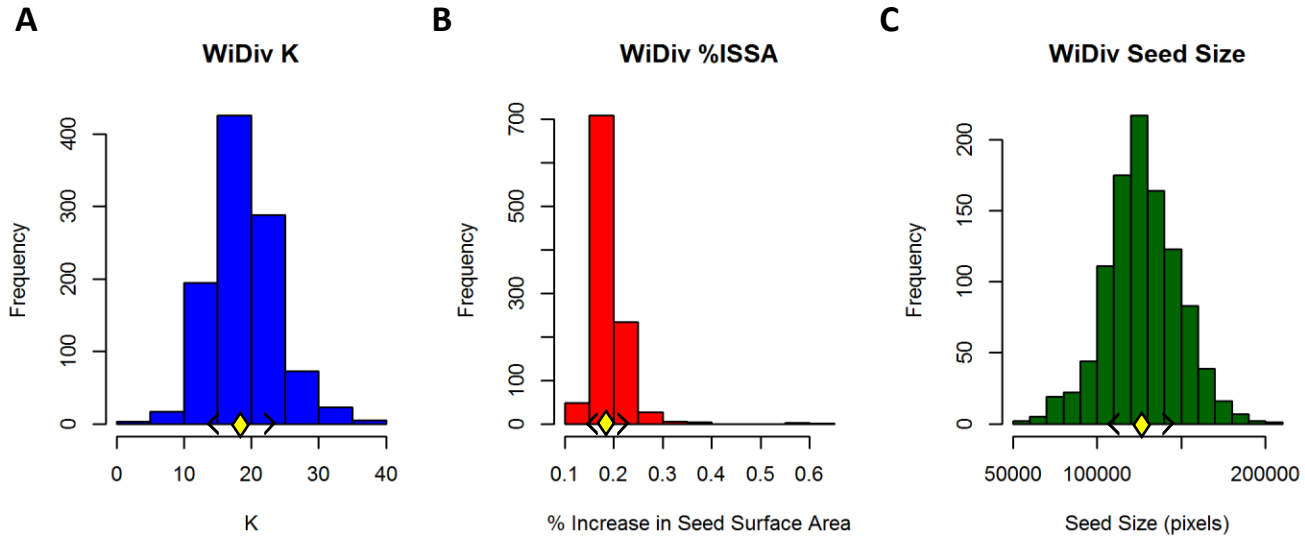


Figure 1. Histograms depicting phenotypic variation for three seed and imbibition traits (*K*, %ISSA, and seed size) measured in the WiDiv population (*n*=500). (A) imbibition rate, given by constant *K*; (B) percentage increase in seed surface area (%ISSA) after 24 hours of imbibition; (C) seed size, given by the seed face surface area in pixels. Yellow triangles and brackets indicate the overall mean and ± one standard deviation from the mean, respectively.

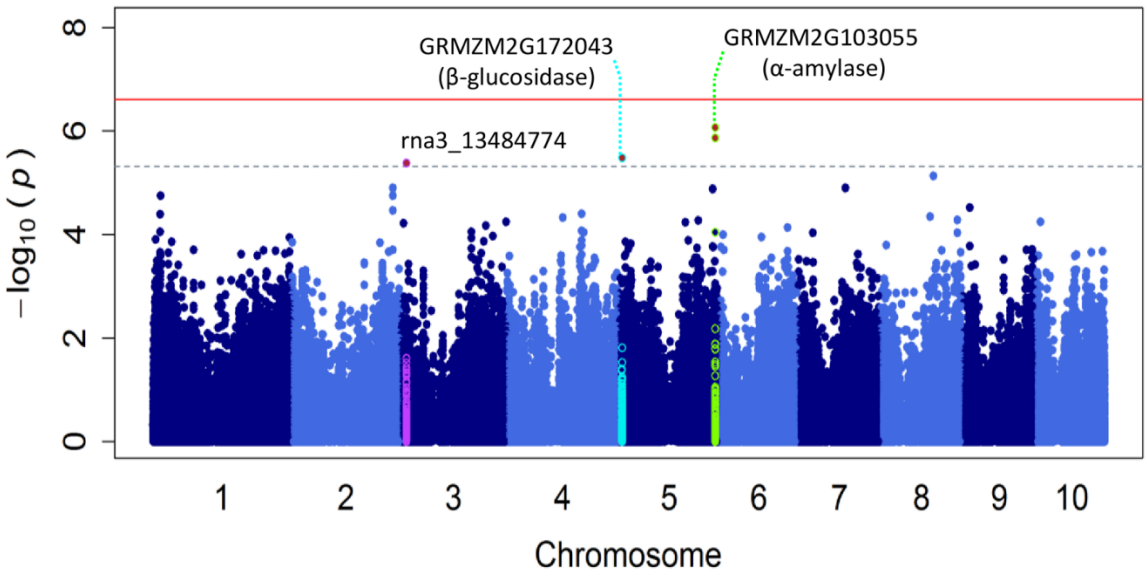


Figure 2. Manhattan plot from GWAS results for maize kernel imbibition rate *K* measured in the WiDiv association panel (*n*=500). The y-axis of the Manhattan plot represents –log10 *p*-values from a genome-wide scan which are plotted against the position of the ~437,000 SNPs of the ten maize chromosomes (depicted on the x-axis). Quantile-quantile (Q-Q) plot depicted in **Appendix A8** indicates model-fit of the GWAS of trait *K*. The solid red line indicates the Bonferroni-adjusted (simple*M* method) genome-wide significance threshold (P = 2.44 x 10^{-7}), and the dashed grey line indicates the genome-wide suggestive threshold (P = 4.90 x 10^{-6}). Suggestive and significant trait-associative SNPs (TASs) are colored red with the functional annotations of the corresponding candidate gene model shown above. SNPs belonging to candidate genes with significant TASs are colored individually on a per candidate gene basis.

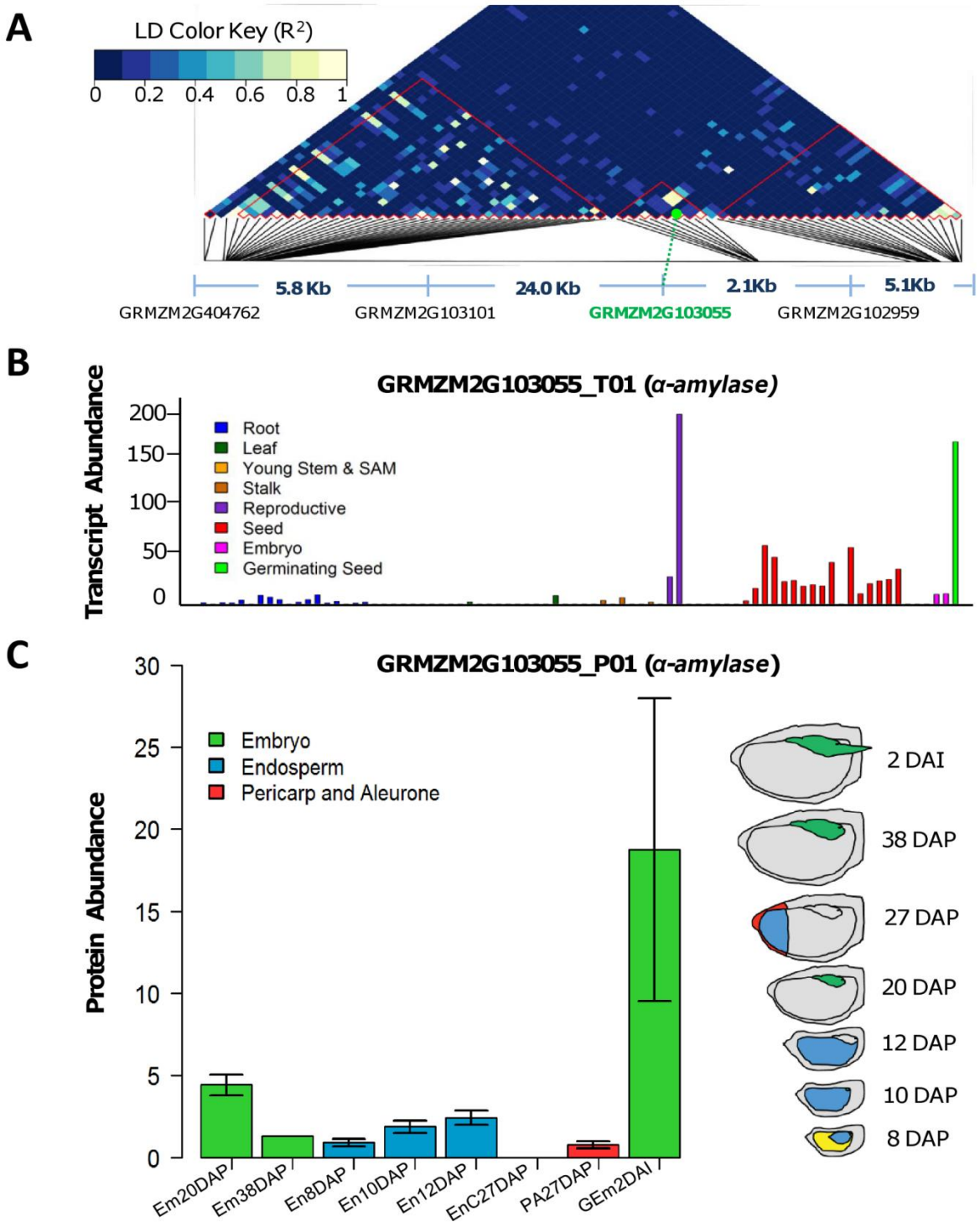

A
LD Color Key (R2)
0 0.2 0.4 0.6 0.8 1
5.8 Kb
24.0 Kb
2.1Kb
5.1Kb
GRMZM2G404762
GRMZM2G103101
GRMZM2G103055
GRMZM2G102959
B
GRMZM2G103055_T01 (α-amylase)
Transcript Abundance
200
150
100
50
0
Root
Leaf
Young Stem & SAM
Stalk
Reproductive
Seed
Embryo
Germinating Seed
C
GRMZM2G103055_P01 (α-amylase)
Protein Abundance
30
25
20
15
10
5
0
Embryo
Endosperm
Pericarp and Aleurone
Em20DAP
Em38DAP
En8DAP
En10DAP
En12DAP
EnC27DAP
PA27DAP
GEm2DAI
2 DAI
38 DAP
27 DAP
20 DAP
12 DAP
10 DAP
8 DAP

Figure 3.

Figure 3. Biological systems multi-omics analysis of a genetic region associated with imbibition rate *K* implicates candidate gene GRMZM2G103055, which encodes a putative *α*-amylase protein. (A) Pairwise linkage disequilibrium (LD; calculated as R^2) for the 37 Kb region on chromosome five surrounding two significant trait-associative SNPs (TASs), with the most significant TAS and its corresponding gene model highlighted in green. Red boundaries indicate pairwise LD between markers within an individual gene model with the gene model IDs listed below the boundaries. Distances between gene models are reported as the physical distance between the ending position of the former gene model and the starting position of the latter. (B) RNA-seq derived gene expression values of candidate gene GRMZM2G103055 from 80 diverse tissues representing 12 organs spanning development of inbred line B73 (STELPFLUG *et al.* 2015). Transcript abundance values are given in fragments per kilobase per millions of reads mapped (FPKM). Individual tissues are colored by organ type as described in legend. (C) Protein abundance of peptide GRMZM2G103055_P01 from eight unique seed tissues spanning grain filling and seed maturation of inbred line B73 (WALLEY *et al.* 2013). Values of protein abundance were derived from mass spectrometry and are given in normalized spectral counts. Error bars denote standard error of normalized spectral counts of 4-7 biological tissue replicates. Cartoon of seed tissues was adapted from Walley *et al.* 2013. Kb: kilobase, Em: embryo, En: endosperm, EnC: endosperm crown, PA: pericarp and aleurone layer, GEm: germinating embryo, DAP: days after pollination, DAI: days after imbibition.

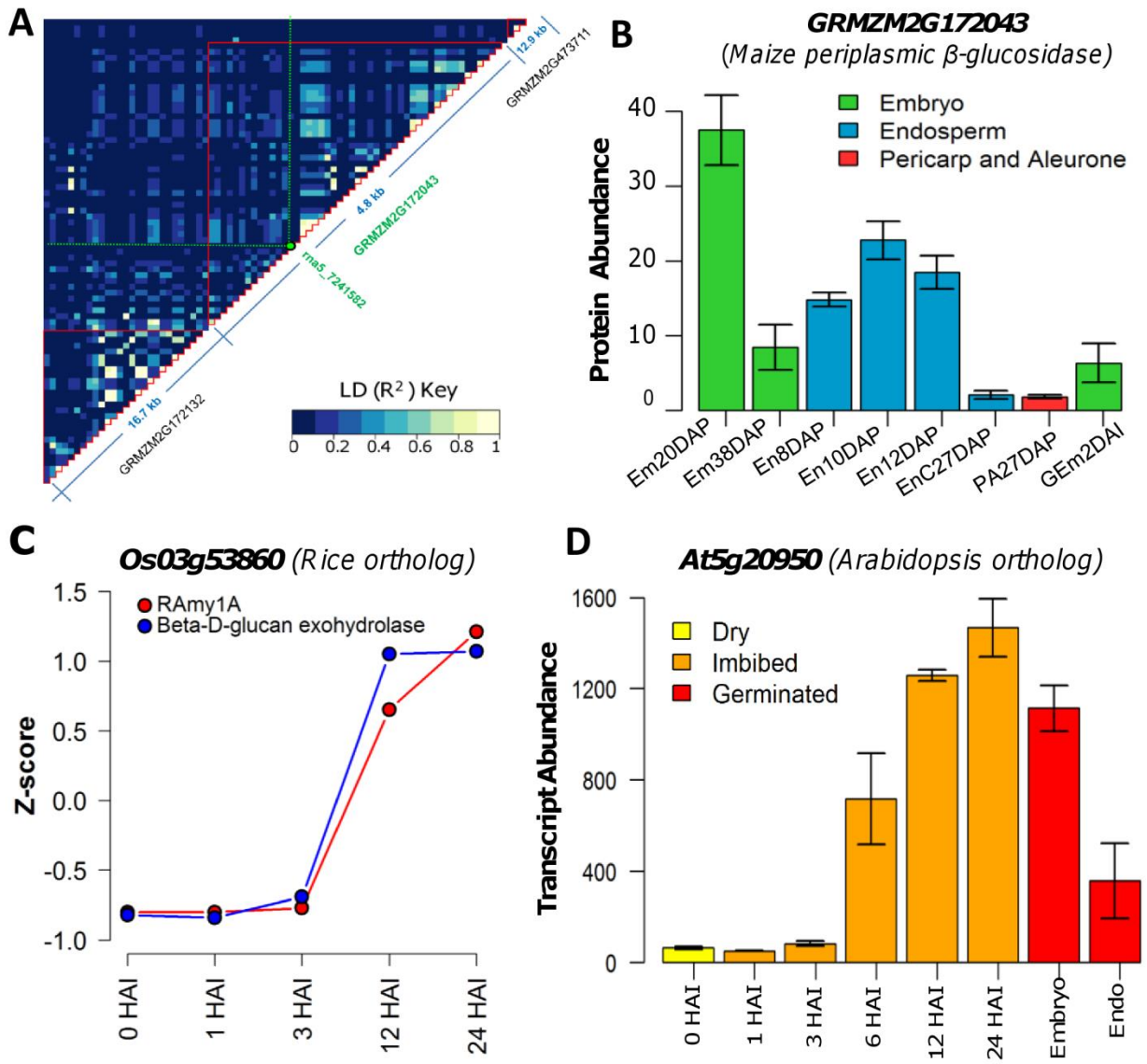

A
12.9 Kb
GRMZM2G473711
4.8 Kb
GRMZM2G172043
rna5_7241582
16.7 Kb
GRMZM2G172132
LD (R²) Key
0 0.2 0.4 0.6 0.8 1
B
GRMZM2G172043
(Maize periplasmic β-glucosidase)
Protein Abundance
Embryo
Endosperm
Pericarp and Aleurone
Em20DAP
Em38DAP
En8DAP
En10DAP
En12DAP
EnC27DAP
PA27DAP
GEm2DAI
C
Os03g53860 (Rice ortholog)
RAmy1A
Beta-D-glucan exohydrolase
Z-score
0 HAI
1 HAI
3 HAI
12 HAI
24 HAI
D
At5g20950 (Arabidopsis ortholog)
Transcript Abundance
Dry
Imbibed
Germinated
0 HAI
1 HAI
3 HAI
6 HAI
12 HAI
24 HAI
Embryo
Endo

Figure 4.

Figure 4. Biological systems multi-omics analysis of a genetic region associated with imbibition rate *K* implicates candidate gene GRMZM2G172043, which encodes a periplasmic β-glucosidase protein. (A) Pairwise linkage disequilibrium (LD; calculated as R^2) for the 34.4 Kb region on chromosome five surrounding TAS rna5_7241582 and its corresponding gene model highlighted in green. Red boundaries indicate pairwise LD between markers within an individual gene model with the gene model IDs listed below the boundaries. Distances between gene models are reported as the physical distance between the ending position of the former gene model and the starting position of the latter. (B) Protein abundance of peptide GRMZM2G172043_P01 from eight unique seed tissues spanning grain filling and seed maturation of inbred line B73 (WALLEY *et al.* 2013). Values of protein abundance were derived from mass spectrometry and are given in normalized spectral counts. Error bars denote standard error of normalized spectral counts of 4-7 biological tissue replicates. (C) Z-score of transcript abundance of rice gene Os03g53860, a syntenic ortholog of maize candidate gene GRMZM2G172043, during early rice germination. Also pictured is the transcript abundance of rice gene *RAmy1A*, the syntenic ortholog of maize candidate gene *ZmAmy1A*, which is co-expressed with rice gene Os03g53860. Expression analysis was interrogated via the Affymetrix Rice Gene Chip microarray of three biological tissue replicates (HOWELL *et al.* 2009). (D) Gene expression profiling of *Arabidopsis* gene At5g20950, a syntentic ortholog of maize candidate gene GRMZM2G172043 during imbibition and germination using the Affymetrix ATH1 Gene Chip microarray (NAKABAYASHI *et al.* 2005). Kb: kilobase; Em: embryo; En: endosperm; EnC: endosperm crown; PA: pericarp and aleurone layer; GEm: germinating embryo; DAP: days after pollination; DAI: days after imbibition.

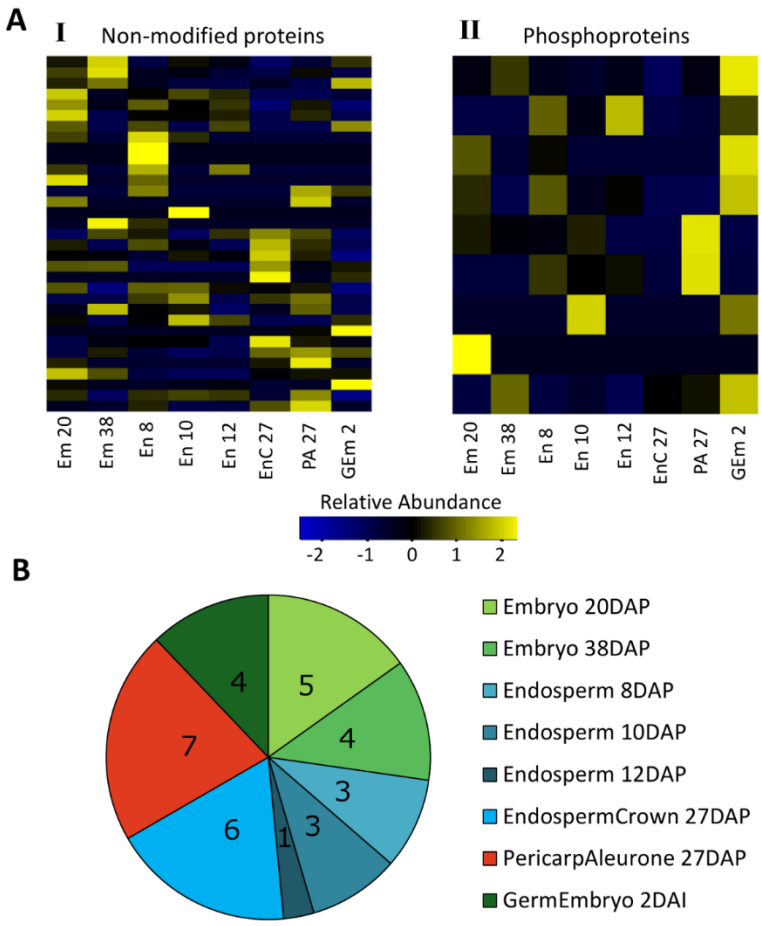


Figure 5. Protein abundance and phosphorylation levels of top GWAS gene hits associated with imbibition rate *K* across unique seed tissues during development and germination. (AI) Hierarchical clustering of non-modified protein abundance of the top 50 GWAS gene hits associated with imbibition rate K across diverse seed tissues spanning seed development (*n* = 33 genes) (AII) Hierarchical clustering of phosphorylation status of the top 50 GWAS gene hits associated with imbibition rate (*n* = 9 genes). Data were extracted from the maize seed proteomic atlas of inbred B73 (WALLEY *et al.* 2013) (B) Pie chart displaying counts of peak non-modified protein abundance of top gene hits across diverse maize seed tissues corresponding to panel AI. Em: embryo; En: endosperm; EnC: endosperm crown; PA: pericarp and aleurone; GEm: germinating embryo; DAP: days after pollination; DAI: days after imbibition.

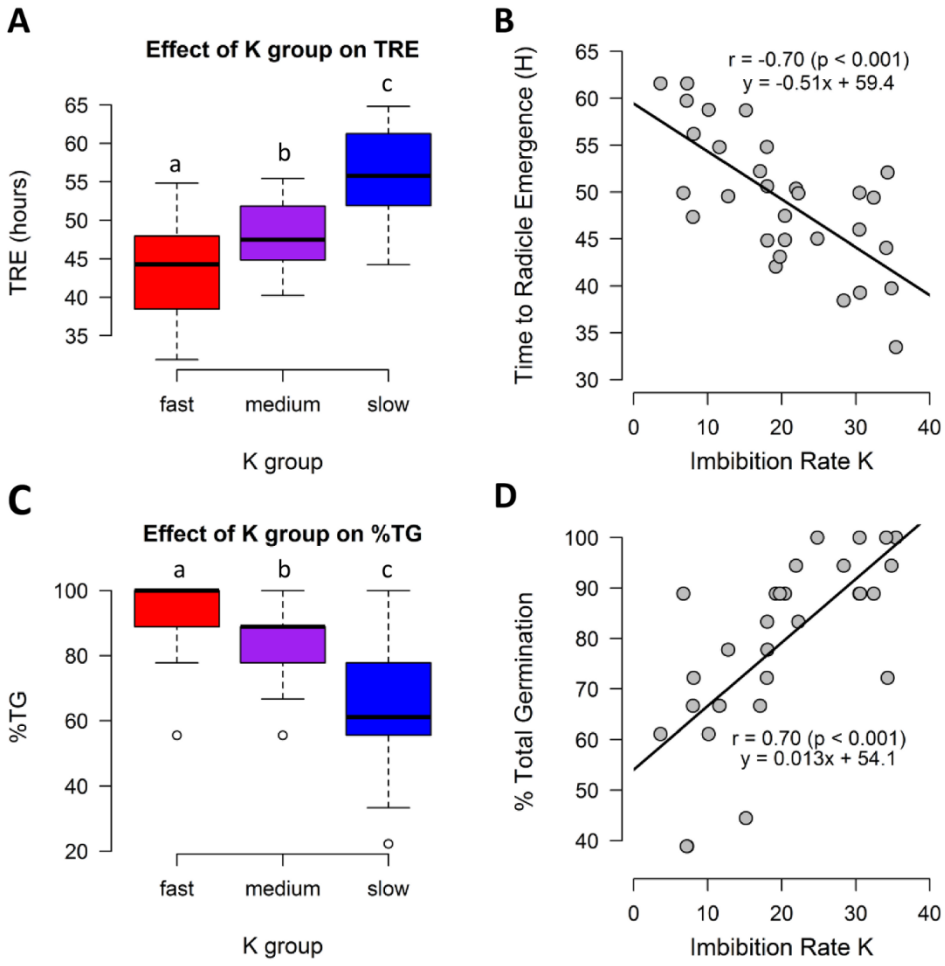


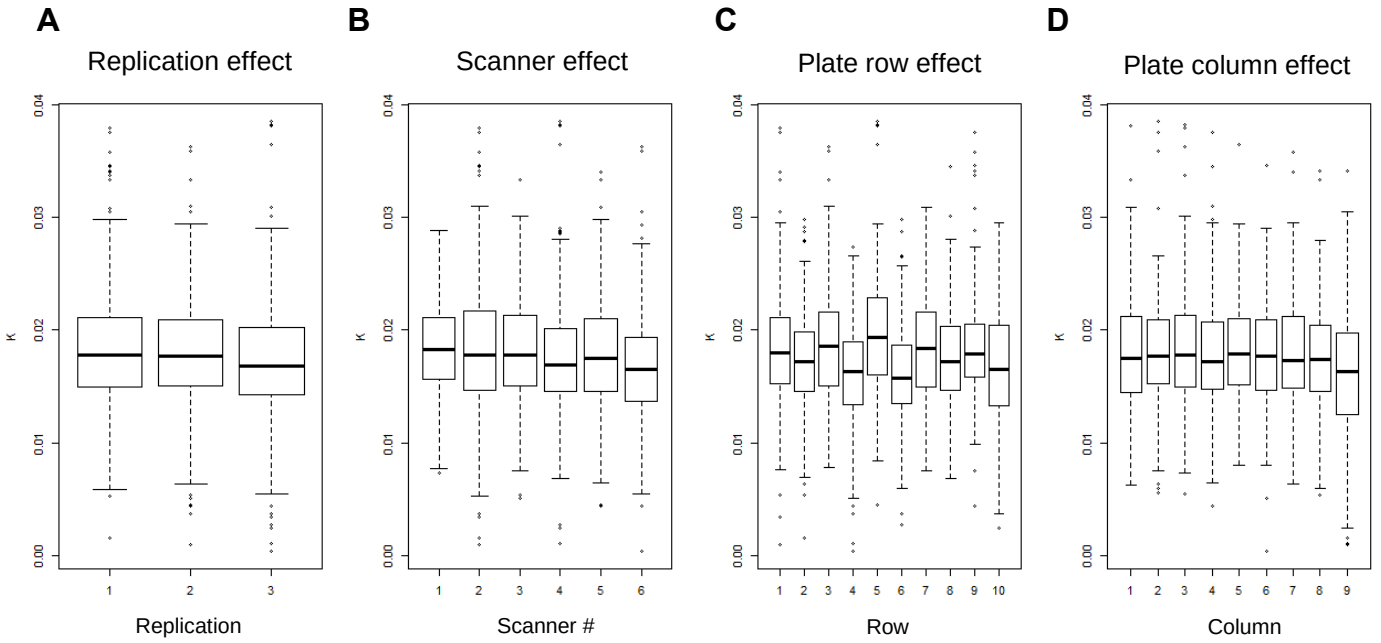
Figure 6. Evaluating the relationship between seed imbibition rate *K* and germination traits using 30 extremely divergent genotypes sampled from the WiDiv association panel. Three *K* groups termed "fast," medium," and "slow" were formed to include the ten fastest, ten slowest, and ten most intermediate imbibing genotypes from the WiDiv panel (listed in **Appendix A5**). Box-plots depict mean differences in (A) time to radicle emergence in hours (TRE) and (C) total germination percentage (%TG) between *K* groups. Different letters (a,b,c) above box-plots indicate statistically significant differences in means based on pair-wise t-tests. Linear plots of Pearson's correlations are shown between (B) *K* vs. TRE and (D) *K* vs. %TG.

APPENDICES

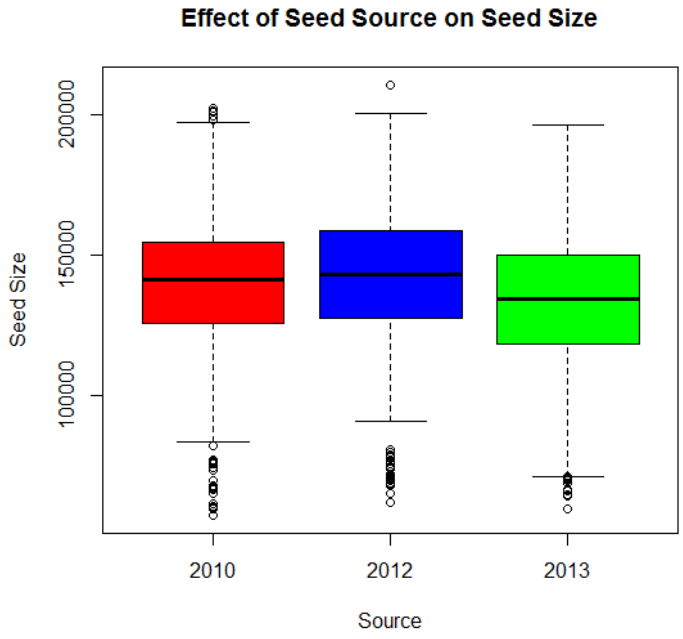
Appendix A1. Pedigree, heterotic groupings, release year and origin information of inbred lines used as parents of the full-diallel experiment and inbreds used for the seed source experiment. Yellow colored cells indicate inbred lines used in both the full-diallel and seed source experiments. SSS: stiff-stalk synthetic; NSS: non-stiff stalk; POP: popcorn; PVP: Plant Variety Protection lines; LH: Holden's

Genotype	Pedigree	Heterotic Group	Year	Origin	PVP
A663	A427(3)X Cuzco	unclassified	1975	U of Minnesota	non
B14A	Cuzco X B14 (8)	SSS	1962	Iowa	non
B73	BSSS	SSS	1972	Iowa	non
CH157	NA	unclassified	.	Ontario, Canada	non
DKHBA1	PH3195 X PH3199	unclassified	1986	United States	PVP-Dekalb
H121	Syn. HCBSA	unclassified	.	Purdue	non
H99	Illinois Syn. 60C	NSS	1974	Indiana	non
IDS69	NA	POP	.	Iowa	non
Ky226*	NCIaDDC X Coah. 8	unclassified	1966	Kentucky	non
LH145	A632 X CM105	SSS	1984	Iowa	PVP-LH
Mo17	C.I.187-2 X C103	NSS	1966	Missouri	non
Mo5	K55(3) X N6	unclassified	1958	Missouri	non
MoG	Mastadon(Variety from Pennsylvania)	unclassified	.	.	non
NC230	K55 X Yellow line or hybrid	unclassified	.	North Carolina	non
ND262	NDSF	unclassified	.	N Dakota	non
NK740	Mexican_Deep_Kernel X Mo17(4)	unclassified	1988	Minnesota	PVP-NTP
Oh43	W8 X Oh40B	NSS	1949	Ohio	non
PH207	PHG3BD2 X PHG3RZ1	NSS (iodent)	1984	Iowa	PVP- Pioneer
PHB47	SD105 X B37(3)	SSS	1984	Iowa	PVP- Pioneer
PHW65	PH861 X PH595	NSS	1988	Iowa	PVP- Pioneer
PHZ51	PH814 X PH848	NSS	1986	Iowa	PVP- Pioneer
SD40	P3709	unclassified	.	S Dakota	non
Tr	Reid Yellow Dent (Troyer strain)	unclassified	1938	Purdue	non
Va99	Oh07B X Pa91	unclassified	.	Virginia	non
W22	Ill_B10 X W25	NSS	1946	Wisconsin	non
WU-TAN-TZAO	NA	unclassified	.	Shaanxi, China	non

***** indicates the line was used only in the full-diallel experiment and not seed source



Appendix A2. Evaluation of putative image acquisition biases on imbibition rate constant (*K*) for 25 diverse maize inbred lines. (A) Box-plot depicting the variation and mean effect of replication (*n*=3) on *K*. (B) Box-plot depicting the variation and mean effect of scanner (*n*=6) on *K*. (C) Box-plot depicting the variation and mean effect of row positioning of seeds placed on agar plates (*n*=10) relative to *K*. (D) Box-plot depicting the variation and mean effect of column positioning of seeds placed on agar plates.



Appendix A3. Effect of seed source on the phenotypic distributions of seed size for 25 diverse inbreds sampled from the WiDiv population across three maternal environments. Different seed sources correspond to seed produced from different summers of inbred increase during the years 2010, 2012, and 2013. Seed size is given by numbers of pixels.

Appendix A4. **List of 596 genotypes used to make the imputed ~437k RNA-seq SNP genotypic dataset for GWAS.** Genotypes having phenotypic data which were included in the GWAS are marked with an "X," as are the 503 genotypes which were included in the previous WiDiv genotypic set (HIRSCH *et al.* 2014). Group indicates heterotic grouping, year indicates year of release, and origin is the site of the breeding station which released the line. NSS: non-stiff stalk; SSS: stiff-stalk synthetic; POP: popcorn; SWEET: sweet corn; TROP: tropical.

Genotype	Swelling Data?	In Hirsch et al. 2014?	Group	Year	Origin
04033V			.	.	.
2FACC			.	1990	United States
33-16			.	1938	Purdue
38-11			.	1938	Purdue
4226		X	NSS	1938	U of Illinois
4554 INBRED		X	.	.	Hungary
4578 INBRED		X	.	.	Hungary
4722			POP	.	Purdue
4F-35 BK		X	.	.	.
4N506			SSS	1990	Illinois
52220			.	.	U of Peking
6M502			.	1990	United States
80-2		X	.	.	Poland
A	X	X	NSS	1939	U of Illinois
A12	X	X	NSS	1946	U of Minnesota
A15	X	X	NSS	1946	U of Minnesota
A155	X	X	.	.	U of Minnesota
A171	X	X	.	1946	U of Minnesota
A188	X	X	.	1946	U of Minnesota
A208	X	X	.	1955	U of Minnesota
A239	X	X	NSS	1957	U of Minnesota
A258	X	X	NSS	.	U of Minnesota
A305	X	X	.	1946	U of Minnesota
A321	X	X	.	.	U of Minnesota
A322	X	X	.	1946	U of Minnesota
A334	X	X	.	1946	U of Minnesota
A340		X	.	1946	U of Minnesota
A374	X	X	.	1946	U of Minnesota

A385	X	X	.	1946	U of Minnesota
A401	X	X	.	1955	U of Minnesota
A415-1-3 INBRED	X	X	.	.	Transvaal, South Africa
A427	X	X	.	1956	U of Minnesota
A554	X	X	NSS	.	U of Minnesota
A556	X		.	1956	Minnesota
A572	X	X	.	.	U of Minnesota
A627	X	X	.	.	U of Minnesota
A632	X	X	SSS	1964	U of Minnesota
A634	X		.	1966	Minnesota
A635	X	X	SSS	1966	U of Minnesota
A641	X		.	1966	Minnesota
A648	X	X	NSS	.	U of Minnesota
A649	X	X	.	1966	U of Minnesota
A651	X	X	.	1966	U of Minnesota
A659	X	X	NSS	.	U of Minnesota
A661		X	.	1975	U of Minnesota
A662	X	X	.	1975	U of Minnesota
A663	X	X	.	1975	U of Minnesota
A673	X	X	.	.	U of Minnesota
A674	X	X	.	.	U of Minnesota
A679	X	X	SSS	.	U of Minnesota
A680	X	X	SSS	.	U of Minnesota
A682	X	X	NSS	.	U of Minnesota
A71	X	X	NSS	1946	U of Minnesota
A73	X	X	.	1946	U of Minnesota
A797NW	X	X	.	.	KwaZulu-Natal, South Africa
AR228	X	X	.	.	Arkansas
AS5707	X	X	.	.	Michigan
AusTRCF 306238	X	X	.	.	NSW, Australia
B 18_INBR FR SUPERGOLD			POP	.	Iowa
B-28		X	POP	.	Iowa
B10	X	X	SSS	1953	Iowa
B101	X	X	SSS	.	Iowa
B103	X	X	.	.	Iowa
B104	X	X	SSS	.	Iowa
B105	X	X	SSS	.	Iowa
B106	X	X	SSS	.	Iowa
B107	X	X	.	.	Iowa
B108	X	X	.	.	Iowa

B109	X	X	SSS	.	Iowa
B110	X	X	SSS	.	.
B111	X	X	SSS	.	.
B112	X	X	.	.	.
B113	X	X	NSS	.	.
B114	X	X	.	.	.
B115	X	X	NSS	.	.
B119	X	X	.	.	.
B120	X	X	SSS	.	.
B121	X	X	SSS	.	.
B14	X	X	SSS	1953	Iowa
B14A	X	X	SSS	1962	Iowa
B164	X	X	.	.	Minnesota
B2	X	X	NSS	1936	Purdue
B37	X	X	SSS	1958	Iowa
B42		X	SSS	1960	Iowa
B46	X		.	.	Iowa
B47	X		.	1960	.
B52		X	NSS	.	Iowa
B54		X	SSS	.	Iowa
B64	X	X	SSS	1966	Iowa
B65			.	1966	Iowa
B66	X	X	.	.	Iowa
B68	X	X	SSS	.	Iowa
B7	X	X	.	1953	Iowa
B73	X	X	SSS	1972	Iowa
B75	X	X	.	1975	Iowa
B76	X		.	1973	Iowa
B77	X	X	NSS	1974	Iowa
B79	X	X	.	1975	Iowa
B8		X	.	1948	Iowa
B84	X	X	SSS	1978	Iowa
B85	X	X	.	1978	Iowa
B87	X	X	.	1981	Iowa
B88	X	X	.	.	Iowa
B90	X	X	.	.	Iowa
B91	X	X	.	.	Iowa
B97	X	X	.	.	Iowa
B98	X	X	NSS	.	Iowa
B99	X	X	.	.	Iowa
C102	X	X	.	1946	Connecticut

C103	X	X	NSS	1946	Connecticut
C123	X	X	NSS	.	Connecticut
C15	X	X	SWEET	1946	Connecticut
C42	X	X	NSS	1946	Minnesota
C49A	X	X	NSS	.	Minnesota
C68	X	X	.	.	Connecticut
CG10	X	X	.	.	Ontario, Canada
CG106		X	.	.	.
CG108	X	X	.	.	.
CG65	X	X	.	.	Ontario, Canada
CH157	X	X	.	.	Ontario, Canada
CH701-30	X	X	.	.	Ontario, Canada
CH753-4	X	X	.	.	Ontario, Canada
CH9	X	X	.	.	Ontario, Canada
CHI-TAN 120		X	.	.	Jilin, China
CI 40H	X	X	.	.	Missouri
CI 187-2		X	.	.	Missouri
CI 21E	X	X	NSS	1952	United States
CI 28A	X	X	.	1956	United States
CI 31A			.	.	.
CI 3A		X	.	1946	Missouri
CI 64	X	X	.	1955	United States
CI 90C	X	X	.	.	United States
CI 91B	X	X	.	1958	United States
CL17	X	X	.	.	Ontario, Canada
CL18	X	X	.	.	Ontario, Canada
CL22		X	.	.	.
CL27	X	X	.	.	Ontario, Canada
CM105		X	SSS	.	Manitoba, Canada
CM174	X	X	SSS	.	Manitoba, Canada
CM37	X		NSS	.	Manitoba, Canada
CM48	X	X	.	.	Manitoba, Canada
CM7	X	X	NSS	.	Manitoba, Canada
CM99	X	X	SSS	.	Manitoba, Canada
CML 108		X	.	.	Federal District, Mexico
CML 144		X	TROP	.	.
CML 161		X	TROP	.	.
CML 176		X	TROP	.	.
CML 197		X	TROP	.	.
CML 202		X	TROP	.	.
CML 216		X	TROP	.	.

CML 218	X	X	TROP	.	Federal District, Mexico
CML 220	X	X	TROP	.	Federal District, Mexico
CML 228	X	X	TROP	.	Federal District, Mexico
CML 247	X	X	TROP	.	Federal District, Mexico
CML 251		X	TROP	.	.
CML 254		X	TROP	.	Federal District, Mexico
CML 264	X	X	TROP	.	Federal District, Mexico
CML 277		X	TROP	.	Federal District, Mexico
CML 287	X	X	TROP	.	Federal District, Mexico
CML 312		X	TROP	.	.
CML 322	X	X	TROP	.	Federal District, Mexico
CML 323	X	X	TROP	.	Federal District, Mexico
CML 376		X	TROP	.	Federal District, Mexico
CML 421		X	TROP	.	.
CML 444		X	TROP	.	.
CML 448		X	TROP	.	.
CML 451		X	TROP	.	.
CML 456		X	TROP	.	.
CML 465		X	TROP	.	.
CML 488		X	TROP	.	.
CML 491		X	TROP	.	.
CML 505		X	TROP	.	.
CML 509		X	TROP	.	.
CML 511		X	TROP	.	.
CML 52		X	TROP	.	Federal District, Mexico
CML 91	X	X	TROP	.	Federal District, Mexico
CML 387			.	.	.
CO106	X	X	NSS	.	Ontario, Canada
CO117	X	X	.	.	Ontario, Canada
CO125	X	X	NSS	.	Ontario, Canada
CO158	X	X	.	.	Ontario, Canada
CO192	X	X	.	.	Ontario, Canada
CO216	X	X	.	.	Ontario, Canada
CO236	X	X	.	.	Ontario, Canada
CO237	X	X	.	.	.
CO245	X	X	.	.	Ontario, Canada
CO255	X	X	.	.	Ontario, Canada
CO256		X	SSS	.	Ontario, Canada
CO257	X	X	SSS	.	Ontario, Canada
CO258	X	X	SSS	.	Ontario, Canada
CR 22 INBRED	X	X	.	.	.

CR14	X		.	1990	United States
CR1HT	X	X	.	1984	United States
CSJ3	X	X	.	.	Ontario, Canada
DE3	X	X	.	.	.
DE4	X	X	.	.	.
DE811	X	X	NSS	.	Delaware
DK4676A	X		.	1986	United States
DK78002A	X		.	1986	United States
DK78371A	X		.	1988	United States
DKFBHJ	X		SSS	1988	United States
DKHBA1	X		.	1986	United States
DKIB014	X		.	1986	United States
DKIB02	X		.	1988	United States
DKMBNA	X		.	1986	United States
DKMBPM	X		.	1988	United States
DKMFD 13D		X	.	.	United States
DKPB80	X		SSS	1988	United States
E2558W	X	X	NSS	.	North Carolina
E8501	X		.	1990	Minnesota
EAST 028	X	X	.	.	Peking, China
Eng-Li Chih	X	X	.	.	Jilin, China
EP1	X	X	.	.	North Carolina
F115		X	.	.	Puy-de-Dome, France
F2834T	X	X	TROP	.	North Carolina
F431	X	X	.	.	.
F44	X	X	NSS	1955	Florida
F7	X	X	.	.	Puy-de-Dome, France
FC46	X	X	.	.	Puy-de-Dome, France
FR19	X	X	SSS	1980	Illinois
G22 T122	X	X	.	.	Romania
G3 T5a		X	.	.	Romania
G80	X	X	.	1984	Iowa
GE129	X	X	.	1961	Georgia
GE54	X	X	.	1961	Georgia
H105W	X		.	.	Indiana
H110	X	X	.	.	Indiana
H113	X	X	.	.	Purdue
H114	X	X	NSS	.	Indiana
H121	X	X	.	.	Purdue
H122w	X	X	SSS	.	Purdue
H124w	X	X	.	.	Purdue

H14	X	X	NSS	.	Purdue
H42	X		.	.	Purdue
H49	X	X	NSS	1959	Purdue
H5	X	X	NSS	1946	Purdue
H52	X	X	.	1959	Purdue
H71	X	X	.	1964	Purdue
H84	X	X	SSS	1966	Indiana
H91	X	X	SSS	1967	.
H95	X	X	NSS	.	Indiana
H96	X	X	.	.	Purdue
H99		X	NSS	1974	Indiana
Hi26	X	X	.	.	Hawaii
Hi28	X	X	.	.	Hawaii
HP301	X	X	POP	.	Indiana
HP72-11	X	X	POP	.	Indiana
HUANYAO	X	X	.	.	Peking, China
HUOBAI		X	.	.	Peking, China
Hy	X	X	NSS	1938	.
I159	X	X	.	.	Iowa
I205	X	X	.	1938	.
I29	X	X	POP	.	Purdue
Ia 453	X	X	POP	.	Iowa
Ia5125B	X	X	POP	.	Iowa
IB02	X	X	.	1988	United States
IBB15	X		.	1990	United States
IBC2	X		.	1990	United States
IDS28		X	POP	.	Iowa
IDS69	X	X	POP	.	Iowa
IDS91	X		POP	.	Iowa
Il 101T	X	X	SWEET	.	.
Il14H	X	X	SWEET	.	.
Il778d	X	X	SWEET	.	.
INB101LFY LFY A632XM16S5	X		.	.	.
INBRED 100		X	.	.	Australia
INBRED 109	X	X	.	.	Australia
INBRED 141	X	X	.	.	Australia
INBRED 305	X	X	.	.	Uruguay
INBRED 309	X	X	.	.	Uruguay
J8606	X		.	.	Minnesota
K148		X	.	1950	Kansas
K150	X	X	.	1950	Kansas

K155	X	X	.	1946	Kansas
K201		X	.	1946	Kansas
K4	X	X	.	1936	Kansas
K41	X	X	NSS	1948	Kansas
K47	X	X	.	.	Kansas
K55	X	X	NSS	1946	Kansas
K64	X	X	NSS	1946	Kansas
Ki11	X	X	SWEET	.	Thailand
Ki21	X	X	.	.	Thailand
Ki3		X	SWEET	.	Thailand
Ki43	X	X	SWEET	.	Thailand
KO679Y	X	X	.	.	South Africa
KUNG-70	X	X	.	.	Shaanxi, China
Ky21	X	X	NSS	1938	Kentucky
Ky226	X	X	.	1966	.
Ky228	X	X	NSS	1966	Kentucky
L 289	X	X	NSS	1938	Iowa
L135	X		.	1990	United States
L139	X		.	1990	United States
L317	X	X	NSS	1946	Iowa
LH1	X	X	.	.	Iowa
LH123HT	X	X	.	1985	Iowa
LH143 (maintainer)	X	X	.	.	Iowa
LH145	X	X	SSS	1984	Iowa
LH146Ht	X		.	.	.
LH149	X	X	SSS	1988	United States
LH156			.	.	.
LH160	X		.	1990	United States
LH38	X	X	.	1981	Iowa
LH39	X	X	NSS	1982	Iowa
LH51	X		.	1983	Iowa
LH59	X	X	.	1987	United States
LH60		X	.	1987	United States
LH61	X	X	.	1987	United States
LH74	X		.	1983	Iowa
LH82	X	X	.	1985	Iowa
LH85			.	1988	United States
LH93	X	X	NSS	1985	Iowa
LP1NRHT	X		.	1979	United States
LP5	X	X	.	1987	France
M14	X	X	NSS	.	U of Illinois

M162W		X	NSS	.	KwaZulu-Natal, South Africa
M37W		X	.	.	KwaZulu-Natal, South Africa
MBPM	X	X	.	1988	United States
MEF156-55-2		X	.	.	North Carolina
Mo15W	X	X	.	.	Missouri
Mo17	X	X	NSS	1966	Missouri
Mo23W	X	X	.	1976	Missouri
Mo24W	X	X	.	1976	Missouri
Mo28W	X	X	.	1976	Missouri
Mo3	X	X	.	1957	Missouri
Mo30W	X	X	.	.	Missouri
Mo39	X	X	.	1976	Missouri
Mo44	X	X	NSS	.	Missouri
Mo45	X	X	.	.	Missouri
Mo46			.	.	Missouri
Mo47	X	X	.	.	Missouri
Mo5	X	X	.	1958	Missouri
Mo7	X	X	.	1959	Missouri
MoG	X	X	.	.	.
Mp339	X	X	.	.	Mississippi
MS106	X	X	.	.	Michigan
MS116	X	X	.	1956	Michigan
MS12	X	X	.	1956	Michigan
MS132	X	X	SSS	.	Michigan
MS142	X	X	.	.	Michigan
MS153	X	X	SSS	1972	Michigan
MS211	X	X	.	1956	Michigan
MS221	X	X	.	.	Michigan
MS222	X	X	.	.	Michigan
MS223	X	X	.	.	Michigan
MS224	X	X	.	.	Michigan
MS225	X	X	.	.	Michigan
MS226	X	X	.	.	Michigan
MS24A	X	X	.	1956	Michigan
MS67	X	X	.	.	Michigan
MS72	X	X	.	.	Michigan
Mt42	X	X	NSS	.	Minnesota
N192	X		.	.	Nebraska
N193	X	X	.	.	Nebraska
N197	X	X	.	.	Nebraska

N199	X	X	.	.	Nebraska
N200	X		.	.	Nebraska
N201	X		.	.	Nebraska
N209		X	.	.	Nebraska
N215	X	X	.	.	Nebraska
N216	X	X	.	.	Nebraska
N217	X	X	SSS	.	Nebraska
N218	X	X	SSS	.	Nebraska
N28	X	X	SSS	.	Nebraska
N28E		X	.	.	.
N28Ht	X	X	SSS	.	.
N501	X	X	SSS	.	Nebraska
N527	X	X	.	.	Nebraska
N534	X	X	.	.	Nebraska
N538	X	X	SSS	.	Nebraska
N542	X	X	SSS	.	Nebraska
N545	X	X	SSS	.	Nebraska
N6	X	X	NSS	1946	Nebraska
N7A	X	X	SSS	.	Nebraska
NC13	X	X	.	1955	North Carolina
NC222	X	X	.	.	North Carolina
NC230		X	.	.	North Carolina
NC232	X	X	.	.	North Carolina
NC236	X	X	.	.	North Carolina
NC250	X	X	SSS	.	North Carolina
NC264	X	X	.	.	North Carolina
NC294			.	.	North Carolina
NC298		X	.	.	North Carolina
NC302	X		.	.	North Carolina
NC306	X		.	.	North Carolina
NC314	X		.	.	North Carolina
NC318	X	X	.	.	North Carolina
NC324	X		.	.	North Carolina
NC326	X		.	.	North Carolina
NC328	X		.	.	North Carolina
NC338	X		.	.	North Carolina
NC340	X	X	.	.	North Carolina
NC342	X	X	.	.	North Carolina
NC344	X	X	.	.	North Carolina
NC348	X		.	.	North Carolina
NC356	X	X	TROP	.	North Carolina

NC358	X	X	TROP	.	North Carolina
NC362	X		.	.	North Carolina
NC364	X		.	.	North Carolina
NC368	X		.	.	North Carolina
NC412	X	X	.	.	.
NC472	X	X	.	.	.
ND1		X	.	1950	N Dakota
ND101		X	.	.	N Dakota
ND167	X	X	.	1948	N Dakota
ND230	X	X	.	1948	N Dakota
ND245	X	X	.	1979	N Dakota
ND246	X	X	NSS	.	N Dakota
ND247	X	X	.	1981	N Dakota
ND251	X	X	.	.	N Dakota
ND260	X	X	.	.	N Dakota
ND262	X	X	.	.	N Dakota
ND265	X	X	.	.	N Dakota
ND283	X	X	.	1948	N Dakota
NK740	X		.	1988	Minnesota
NK778	X		.	.	Minnesota
NK807	X		.	.	Minnesota
NKH8431	X		.	1988	Minnesota
NKS8324	X		SSS	1988	Minnesota
NKS8326	X		.	.	.
NO. 380	X	X	.	.	Former Soviet Union
NP87	X	X	POP	.	Nebraska
NY_159_ Neveh_Yaar	X	X	.	.	Israel
NY6371	X	X	.	.	New York
OC19		X	.	.	Iowa
OH33	X	X	.	.	Ohio
Oh40B	X	X	NSS	1946	Ohio
Oh43	X	X	NSS	1949	Ohio
Oh43E	X		.	.	.
Oh7	X	X	.	1946	Ohio
Oh7B		X	.	1956	Ohio
OQ603	X		.	1989	Illinois
Os420	X	X	NSS	1936	Iowa
Os426	X	X	NSS	1936	Iowa
P39	X	X	.	.	Indiana
Pa392	X	X	.	.	Pennsylvania
Pa405	X	X	.	1972	Pennsylvania

Pa468	X	X	.	.	Pennsylvania
Pa762	X	X	NSS	.	Pennsylvania
Pa778	X	X	NSS	.	Pennsylvania
Pa880	X	X	NSS	.	Pennsylvania
Pa891	X	X	.	.	Pennsylvania
Pa91	X		.	.	Pennsylvania
PH207			.	1984	Iowa
PHB09	X	X	.	.	Iowa
PHB47			.	1984	Iowa
PHBA6	X		.	1992	Iowa
PHG29	X	X	.	1986	Iowa
PHG35	X	X	.	1983	Iowa
PHG39	X	X	.	1983	Iowa
PHG47		X	.	1986	Iowa
PHG50	X	X	.	1983	Iowa
PHG71	X		.	1984	Iowa
PHG72	X	X	.	1986	Iowa
PHG80	X	X	.	1984	Iowa
PHG83			.	1985	Iowa
PHG86	X		.	1988	Iowa
PHJ40	X	X	.	1986	Iowa
PHJ75			.	1990	Iowa
PHK35	X		.	1990	Iowa
PHK42	X		.	1988	Iowa
PHK76	X	X	.	1988	Iowa
PHN11	X	X	.	1988	Iowa
PHP02	X		.	1988	Iowa
PHP55	X		.	1990	Iowa
PHP85	X		.	1992	Iowa
PHPR5	X		.	1992	Iowa
PHR30	X		.	1992	Iowa
PHR36	X		.	1986	Iowa
PHR62	X		.	1990	Iowa
PHT22	X		.	1990	Iowa
PHT55	X		.	1988	Iowa
PHT77	X	X	.	1988	Iowa
PHV37	X		.	1990	Iowa
PHV53	X		.	1992	Iowa
PHV63	X	X	.	1988	Iowa
PHVA9	X		.	1992	Iowa
PHW03	X		.	1990	Iowa

PHW17	X	X	.	1986	Iowa
PHW20	X		.	1990	Iowa
PHW65	X		.	1988	Iowa
PHWG5	X		.	1992	Iowa
PHZ51	X	X	.	1986	Iowa
Q381	X		.	1985	United States
R101	X	X	SSS	1957	Illinois
R113	X	X	SSS	1955	Illinois
R134	X	X	.	1957	Illinois
R168	X	X	NSS	1957	Illinois
R177	X	X	NSS	1957	Illinois
R181B	X	X	NSS	.	Illinois
R197	X	X	.	.	Illinois
R225	X	X	.	.	Illinois
R226		X	.	.	Illinois
R227	X	X	.	.	Illinois
R229	X	X	SSS	.	Illinois
R30	X	X	.	1946	Illinois
R4	X	X	NSS	1938	.
R53	X	X	.	1957	Illinois
R71	X	X	NSS	1955	Illinois
R78	X	X	.	1957	Illinois
RS710	X		.	1990	Minnesota
S 56	X	X	.	.	Poland
SA24	X		POP	.	.
SCRO1			.	.	.
SD101	X	X	.	.	S Dakota
SD102	X	X	.	.	S Dakota
SD107	X	X	.	.	S Dakota
SD15	X	X	.	.	S Dakota
SD40	X	X	.	.	S Dakota
SD42	X	X	.	.	S Dakota
SD44		X	NSS	.	S Dakota
SeagullSeventeen	X		.	1980	United States
SG 30A	X	X	POP	1946	Kansas
Sg1533	X	X	POP	.	Indiana
Sg18		X	POP	1946	Iowa
T141		X	.	.	Romania
T146	X	X	.	.	Romania
T232	X	X	.	.	Tennessee
T234	X	X	NSS	.	Tennessee

T242	X	X	.	.	Romania
T9	X	X	.	.	Romania
Tr	X	X	.	1938	Purdue
Tx303	X	X	.	.	Texas
TZU-CHIAO-HIS-WU-105	X	X	.	.	Shaanxi, China
U 123	X	X	.	.	Poland
U267Y		X	.	.	.
Va102	X	X	NSS	.	Virginia
Va14	X	X	NSS	1975	.
Va22	X		.	1975	Virginia
Va26	X		.	.	Virginia
Va35	X		.	.	Virginia
Va38	X	X	.	1975	Virginia
Va52	X	X	.	1975	Virginia
Va59	X		.	1972	Virginia
Va85	X	X	NSS	1975	Virginia
Va99		X	.	.	Virginia
VaW6	X	X	.	.	Virginia
W 7151	X	X	.	.	Wisconsin
W117Ht	X	X	.	.	Wisconsin
W153R	X	X	NSS	1953	Wisconsin
W182BN	X	X	NSS	.	Wisconsin
W182Ngt			.	.	.
W22	X	X	NSS	1946	Wisconsin
W23		X	.	1946	Wisconsin
W24	X	X	.	1950	Wisconsin
W32	X	X	.	1947	Wisconsin
W37A	X	X	.	1955	Wisconsin
W552	X	X	.	.	Wisconsin
W59E	X	X	.	1954	Wisconsin
W601S	X	X	.	.	Wisconsin
W603S	X	X	.	.	Wisconsin
W604S	X	X	.	.	Wisconsin
W605S	X	X	.	.	Wisconsin
W606S	X	X	.	.	Wisconsin
W607S	X	X	.	.	Wisconsin
W608S	X	X	.	.	Wisconsin
W609S	X	X	.	.	Wisconsin
W610S	X	X	.	.	Wisconsin
W611S	X	X	.	.	Wisconsin
W612S	X	X	.	.	Wisconsin

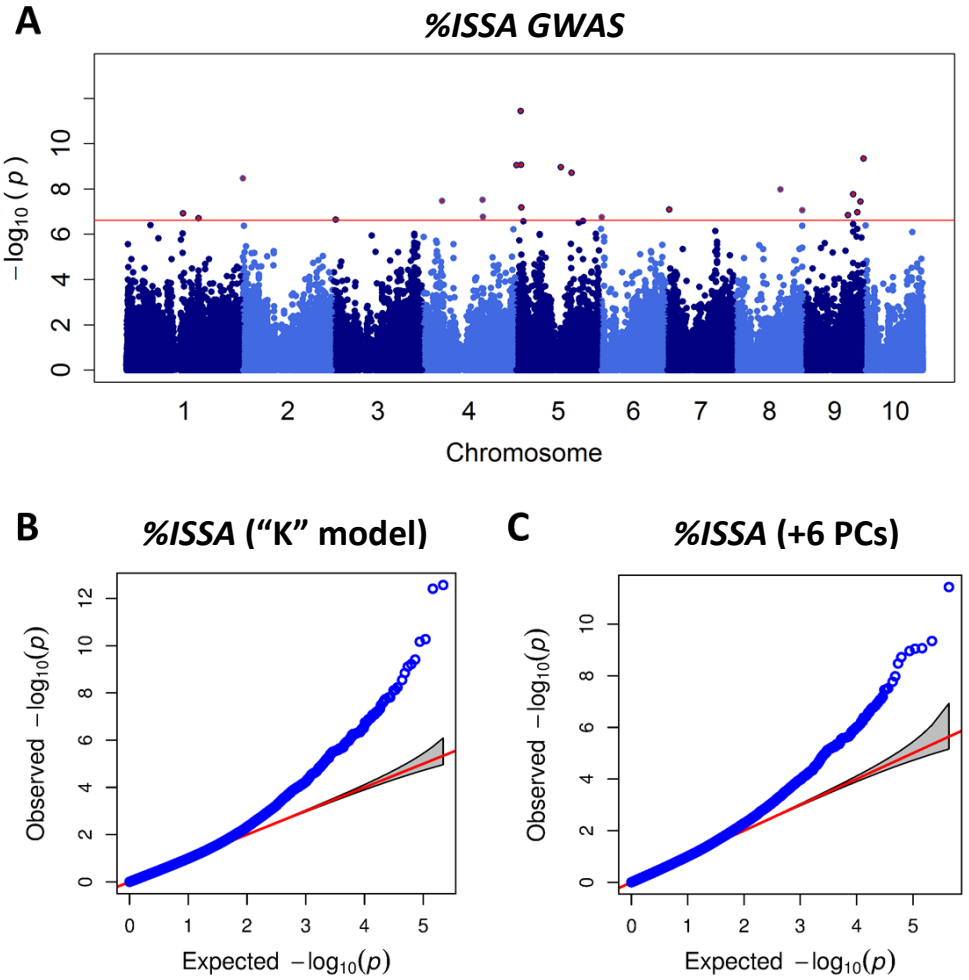
W64A	X	X	NSS	1954	Wisconsin
W701BC	X	X	.	.	Wisconsin
W703	X	X	.	.	Wisconsin
W802G	X	X	.	.	Wisconsin
W803G	X	X	.	.	Wisconsin
W809G	X	X	.	.	Wisconsin
W810G	X	X	.	.	Wisconsin
W811G	X	X	.	.	Wisconsin
W812G		X	.	.	Wisconsin
W813G	X	X	.	.	Wisconsin
W814G	X	X	.	.	Wisconsin
W815G	X	X	.	.	Wisconsin
W816G	X	X	.	.	Wisconsin
W817G	X	X	.	.	Wisconsin
W818G	X	X	.	.	Wisconsin
W819G		X	.	.	Wisconsin
W820G	X	X	.	.	Wisconsin
W821G	X	X	.	.	Wisconsin
W85	X	X	.	1967	.
W8555	X		.	1990	Minnesota
W9	X	X	.	1946	Wisconsin
WD	X	X	.	1946	Wisconsin
WIL500	X		.	1990	United States
WIL900	X		.	1991	United States
WIL901	X		.	1991	United States
WIL903	X		.	1991	United States
WR3	X	X	NSS	1946	Wisconsin
WU-TAN-TZAO	X	X	.	.	Shaanxi, China
WXB6	X	X	.	.	Iowa
YANG	X	X	.	.	Shaanxi, China
YE-CHI-HUNG	X	X	.	.	Shaanxi, China
Ye 4	X	X	.	.	Peking, China
YELLOW 3-4	X	X	.	.	Peking, China
YING-55	X	X	.	.	Shaanxi, China
YONG 28	X	X	.	.	Peking, China
Yu796 NS	X	X	.	.	.
ZS01250	X		.	1996	United States

Appendix A5. Extreme genotypes selected from the WiDiv association panel (*n* = 500) to evaluate the potential relationship between imbibition rate *K* and germination characteristics. Genotypes sampled include the ten fastest, ten slowest, and ten most intermediate genotypes from the WiDiv panel (*n* = 30). Phenotypic values are averaged across both replications of the experiment. *K*: imbibition rate constant; %ISSA: percentage increase in seed surface area after 24 H of imbibition; TRE (H): time to radicle emergence (hours); %TG: percentage of total germination after 72 H.

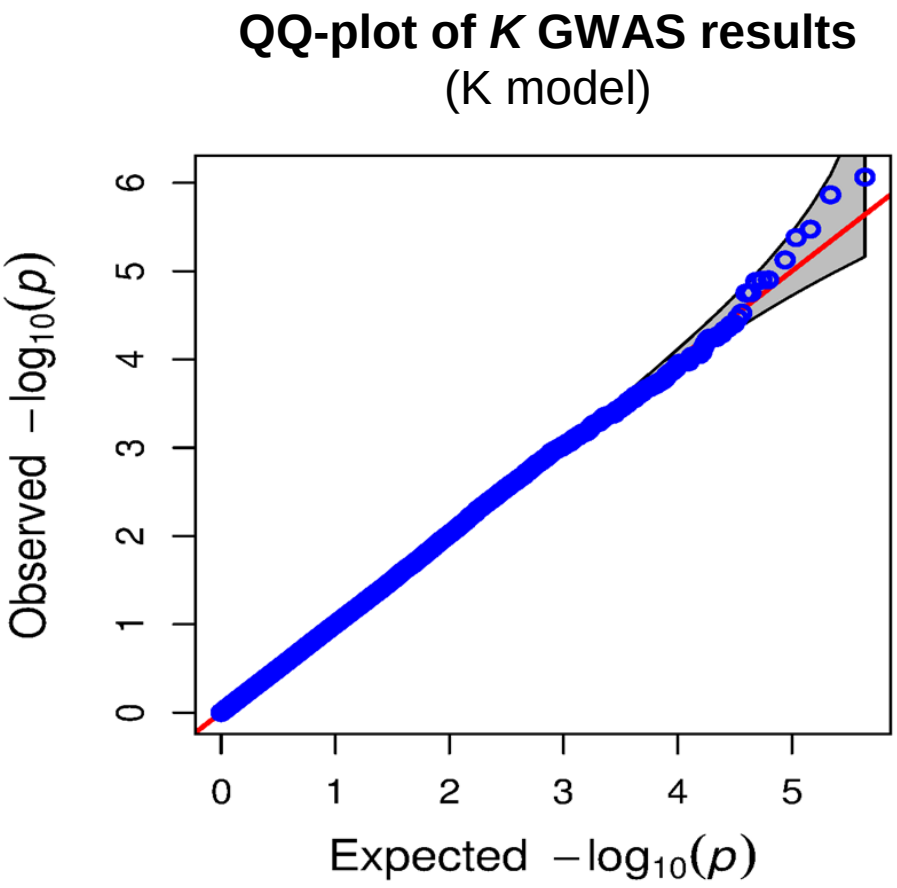
Genotype	Group	*K*	%ISSA	Seed Size	TRE (H)	% TG
WIL900	fast	32.42	22.58%	172294	49.40	88.9%
PHT22	fast	30.47	21.90%	142476	46.00	88.9%
CI90C	fast	35.43	23.45%	72628	33.47	100.0%
Ki43	fast	30.56	22.54%	99405	39.30	88.9%
CML323	fast	34.82	23.39%	96312	39.76	94.4%
N200	fast	34.11	16.91%	145682	44.06	100.0%
NC338	fast	30.49	20.66%	118240	49.92	100.0%
CML322	fast	24.79	19.39%	113043	45.03	100.0%
WIL500	fast	34.30	18.57%	137380	52.09	72.2%
NC356	fast	28.37	21.14%	147800	38.46	94.4%
W64A	medium	18.04	18.04%	130500	44.85	77.8%
CML218	medium	20.4	19.39%	108850	47.46	88.9%
PHG50	medium	17.05	16.62%	152213	52.22	66.7%
SG30A	medium	21.92	17.73%	62731	50.38	94.4%
B109	medium	20.41	16.51%	139005	44.91	88.9%
H124w	medium	19.17	19.10%	148044	42.08	88.9%
SD15	medium	18.01	18.41%	144297	50.63	83.3%
NC326	medium	19.74	19.21%	128798	43.13	88.9%
LH82	medium	22.22	16.57%	132500	49.88	83.3%
L139	medium	17.97	15.51%	164710	54.82	72.2%
Il101T	slow	7.23	38.91%	146241	61.60	38.9%
W7151	slow	8.10	30.23%	115710	56.20	72.2%
P39	slow	7.15	35.60%	140046	59.75	38.9%
CL17	slow	10.11	20.77%	141361	58.78	61.1%
Il14H	slow	3.62	60.04%	161878	61.60	61.1%
CL18	slow	6.70	37.41%	115965	49.91	88.9%
A641	slow	12.73	20.83%	114084	49.55	77.8%
Yong28	slow	15.12	20.60%	128495	58.72	44.4%
C42	slow	8.00	32.87%	171662	47.36	66.7%
EP1	slow	11.57	17.34%	121168	54.80	66.7%

Appendix A6. Pair-wise Pearson's correlations between imbibition rate, percentage increase in seed surface area after 24 H imbibition, and seed size of maize kernels from the WiDiv association panel. Correlations from the seed source and diallel experiments are independent of the WiDiv panel, and are described previously in chapter two. ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, *ns*: not significant ($p > 0.05$); *r*: Pearson's correlation coefficient, *n*: number of kernels phenotyped in the experiment, *K*: imbibition rate constant, %ISSA: percentage increase of seed surface area after 24 hours imbibition.

	r		
Pair-wise Phenotypic Correlation	***Seed Source*** (*n*=1853)	***Diallel*** (*n*=2585)	***WiDiv Association Panel*** (*n*=9212)
K vs. %ISSA	-0.259 ***	-0.434 ***	-0.218 ***
K vs. Seed Size	-0.015 *ns*	0.104 ***	-0.0763 *
%ISSA vs. Seed Size	0.116 ***	0.013 *ns*	0.044 *ns*



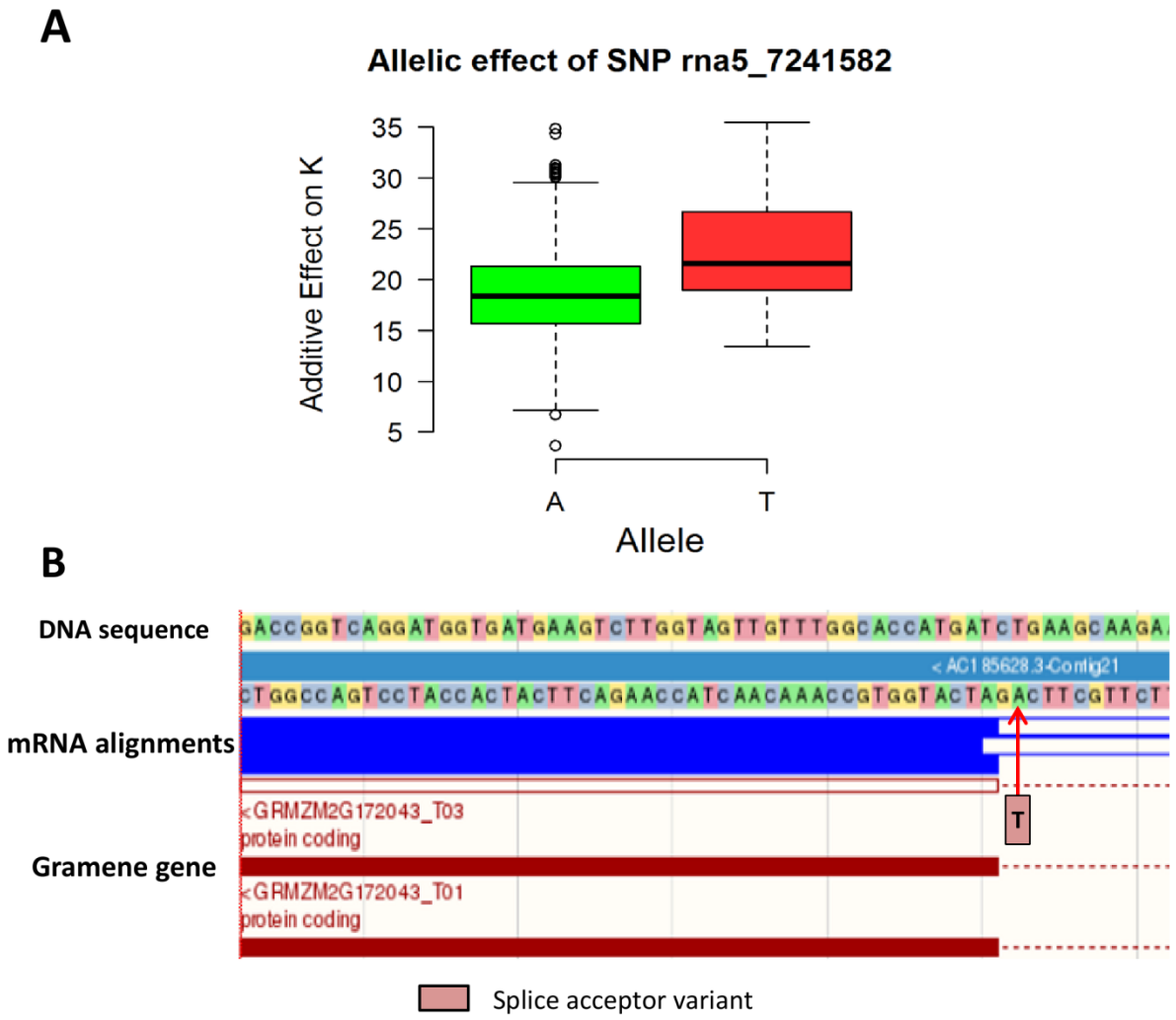
Appendix A7. GWAS results of %ISSA measured in the WiDiv association panel (*n*=500). **(A)** For the Manhattan plot, the y-axis represents –log10 *p*-values from a genome-wide scan which are plotted against the position of ~437,000 SNPs from each of the ten maize chromosomes (depicted on the x-axis). The solid red line indicates the Bonferroni-adjusted (simple*M* method) genome-wide significance threshold (p = 2.44 x 10^{-7}), and the dashed grey line indicates the genome-wide suggestive threshold (p = 4.90 x 10^{-6}). Quantile-Quantile (Q-Q) plots of observed versus expected *p*-values of %ISSA GWAS using the (**A**) "K"model, which includes only the VanRaden kinship matrix to account for the polygenic background and (**B**) the Q+K model, using 6 PCs to account for population structure.



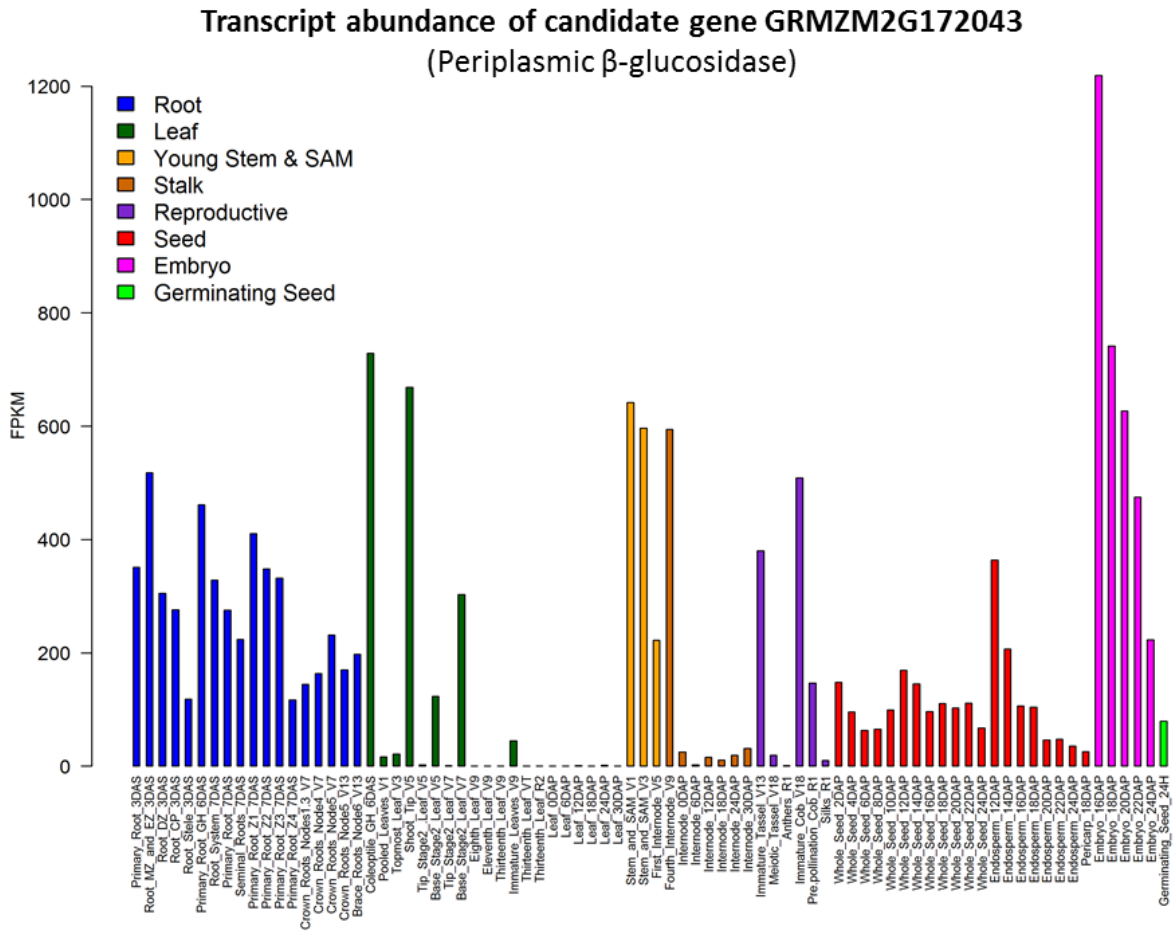
Appendix A8. Quantile-Quantile (Q-Q) plot for observed versus expected *p*-values of imbibition rate *K* GWAS measured in the WiDiv association panel (*n*=500). The "K" model using only the VanRaden kinship matrix to account for the polygenic background effect (VANRADEN 2008). The horizontal axis shows the -log10-transformed expected P-values, and the vertical axis indicates -log10-transformed observed *P*-values.

Appendix A9. Ensembl Variant Effect Prediction (VEP) analysis of trait-associated SNPs with K within Gramene. RefGen_v2 coordinates of the B73 genome were converted to RefGen_v3 coordinates to conduct VEP analysis.

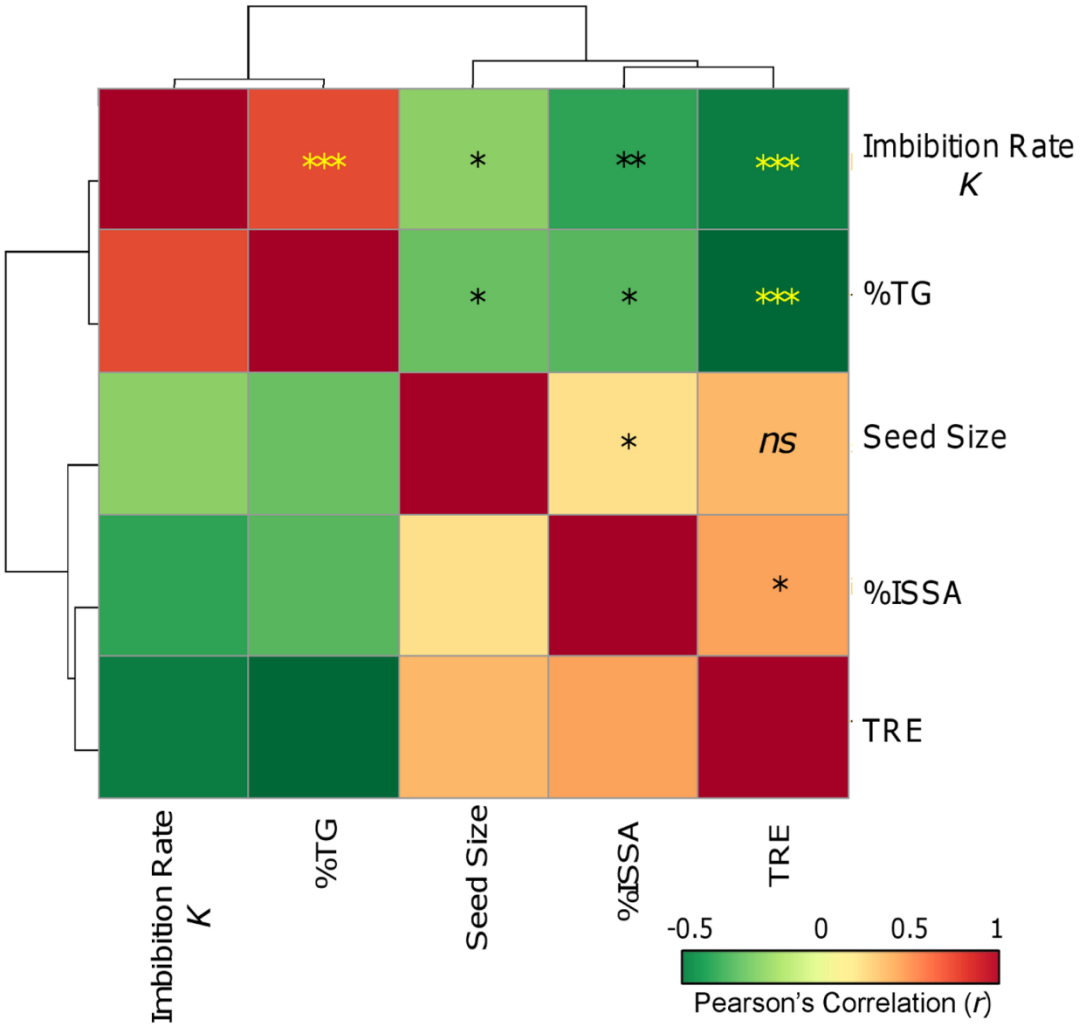
SNP	Location (Chr_Pos)	Allele	Consequence	Impact	Transcript	Exon	Intron	CDS_pos	Amino acid	Codons
A→G	5_209948374	G	synonymous variant	LOW	GRMZM2G103055_T01	2 of 4	-	1041	Alanine	gcA/gcG
T→C	5_209948380	C	synonymous variant	LOW	GRMZM2G103055_T01	2 of 4	-	1047	Alanine	gcT/gcC
A→T	5_7249572	T	splice acceptor variant	HIGH	GRMZM2G172043_T02	-	1 of 5	-	-	-
A→T	5_7249572	T	splice acceptor variant	HIGH	GRMZM2G172043_T01	-	1 of 5	-	-	-



Appendix A10. Allelic characterization of TAS rna5_7241582 within gene GRMZM2G172043, a periplasmic β-glucosidase protein. (A) Effect of allelic substitution from the major allele (A) to the minor allele (T) on imbibition rate *K* from GWAS. (B) Track displaying the DNA sequence, mRNA alignments, and gene model from Gramene via VEP analysis.



Appendix A11. RNA-seq derived gene expression values of candidate gene GRMZM2G103055, a putative periplasmic β-glucosidase protein. Gene expression values are represented from 80 diverse tissues representing 12 organs spanning development of inbred line B73 (STELPFLUG *et al.* 2015). Transcript abundance values are given in fragments per kilobase per millions of reads mapped (FPKM) on the y-axis. Individual tissues are listed on the x-axis and are colored by organ type as described in legend. Tissue names are described in further detail by Stelpflug *et al.* 2015



Appendix A12. Heat map of pairwise Pearson's correlations between all seed, imbibition, and germination traits between the 30 selected extreme genotypes from the WiDiv association panel. %ISSA: percentage increase in seed surface area after 24 H; %TG: percentage total germination after 72 H; TRE: time to radicle emergence (H); ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$; *ns*: not significant ($p > 0.05$).