Genetic architecture and predictability of seedling root traits in maize *(Zea mays* L.*)*

by

Jordon Michael Pace

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Genetics

Program of Study Committee:
Thomas Lübberstedt, Major Professor
Jack Dekkers
Candice Gardner
Michael Lee
Daniel Nettleton
Michael Castellano

Iowa State University

Ames, Iowa

2015

TABLE OF CONTENTS

CHAPTER ONE

GENERAL INTRODUCTION

The maize (*Zea mays* L.) root system is responsible for plant stability and uptake of water and nutrients such as nitrogen (N) in an efficient manner (Lynch, 1995; Aiken and Smucker, 1996). The root system interacts with the rhizosphere (Bais et al., 2006; Watt et al., 2006) and is able to adapt to changing environmental conditions such as excess water, drought conditions, and low nutrient availability (Hawes et al., 1998; McCully, 1988; Drew, 1975). Maize roots are formed during both embryonic and post-embryonic development (Feldman, 1994). There are five main types of roots in maize: crown, seminal, primary, lateral, and brace roots (Hochholdinger, 2009). The major portion of root biomass of mature plants is derived from postembryonic, shoot-borne roots. These postembryonic roots include crown roots and lateral roots, both formed below soil surface, and brace roots, formed above the soil surface (Hoppe et al., 1986). Their function is vital to whole plant performance as they are responsible for the majority of water and nutrient uptake in maize in later developmental stages. Embryonic roots consist of primary and seminal roots. The seminal roots are formed at the scutellar node in the embryo with the number of seminal roots largely dependent on the genetic background of the plant. The overall fate of embryonic roots is background dependent as in some inbred lines; both the primary and seminal roots remain intact in form and function while in other genetic backgrounds these root types become obsolete when compared to postembryonic shoot borne roots (Feldman, 1994).

Two to three week old seedling root systems are made up of primary roots, lateral roots, seminal roots, crown roots and root hairs (Zhu et al., 2006; Hochholdinger, 2009). Lateral roots

are initiated from the pericycle of other roots and have a strong influence on maize root architecture (Esau, 1965, Lynch, 1995). Their function is important to plant performance as they are responsible for a crucial part of water and nutrient uptake in maize (McCully and Canny, 1988). Lateral roots branch outward from the primary root. Lateral roots increase the surface area of the root system and all root types contribute to water and nutrient uptake (W.R. Jordan, 1983; Lynch, 1995; Liu et al., 2008). Moreover, lateral roots contain root initiation points, leading to secondary, tertiary, and higher order root structures, with major influence on the overall root architecture of the root stock (Lynch, 1995).

Genetic control of root development

Understanding of all genetic factors contributing to root development is incomplete. Several genes have been identified that affect root development in maize including *Rtcs* (rootless concerning crown and seminal roots), *Rth1* (roothairless 1), *Rth3* (roothairless 3), and *Rum1* (rootless with undetectable meristems 1). *Rtcs* controls crown root and seminal root formation; *Rtcn* and *Rtcl* are paralogs of *Rtcs. Rth1* and *Rth3* control root hair elongation and development in maize. *Rth3* has been shown to affect adult plant performance for grain yield in maize (Taramino et al., 2007). *Rum1* controls lateral root growth and seminal root growth. A paralog of *Rum1* is *Rul1* (Taramino et al., 2007; Von Behrens et al., 2011). While these genes have been validated through mutational analysis, there are still many loci throughout the genome yet to be discovered that effect root development. To identify additional loci effecting root development, multiple genetic mapping techniques have been developed. Association mapping or linkage disequilibrium (LD) mapping is a useful tool for analyzing the genetic diversity of complex traits and identification of superior alleles (Yan et al., 2011). Unlike traditional linkage mapping, where bi-parental populations are developed, association mapping exploits ancestral

recombination in admixed populations to find marker-trait associations based on LD (Thornsberry, 2002). Researchers have utilized two strategies when conduction association mapping: a candidate gene approach, and a genome-wide approach. Candidate gene association mapping focuses on polymorphisms in predetermined genes known to affect traits of interest, conversely genome-wide association approaches survey the entire genome for polymorphisms associated with complex traits without a priori information of specific regions of the genome that affect traits under study (Merkangas, 1996). The advent of more economic sequencing technologies allowed genome-wide studies as well as genomic selection studies to become more manageable (Metzker, 2010). The type of polymorphism or difference in loci used in genome wide association analysis and found in most abundance within the maize genome are single nucleotide polymorphisms (SNPs) (Rafalski, 2002). SNPs are single nucleotide (A, T, C, or G) variations within the genome that differ between members of a biological species. In candidate gene or genome wide association studies, the relationship between genotypes at a particular locus and phenotype are analyzed to detect whether or not a SNP is statistically associated with a particular trait. Most likely, if found associated, a SNP is in LD with a QTL or gene that affects the trait of interest being studied.

Importance or root development on plant performance

Plants can increase nutrient acquisition by developing larger root systems that will allow contact with a larger soil volume (Tian et al., 2006). Changes in root architecture may have played a large role in adaptation of maize hybrids to today's planting densities, mirroring above ground canopy by narrowing the root branch angles helped for adaptation to higher planting densities (Hammer et al., 2009). Root structure and development have been shown to be a key component to nitrogen use efficiency (NUE) (Hirel et al., 2007) and drought tolerance (Ribaut et

al., 2007). A typical measure for (genetic) NUE in maize is the percentage of grain yield reduction under low N levels compared to high N levels (Presterl et al., 2003).

There is extensive phenotypic variation for root architecture. However, root traits are not routinely directly selected upon by plant breeders for improved nutrient uptake efficiency or yield improvement due to the difficulty in measuring root traits and their quantitative mode of inheritance (Salvi, 2007). Changes in maize root architecture may strongly affect yield (Hammer et al., 2009). Seminal roots play a key role in the acquisition of immobile and mobile nutrients such as phosphorus and N and can determine spatial and temporal domains of its environment and inter-root competition (Zhu et al., 2006). Studying adult roots using "shovelomics", a high-throughput phenotyping technique that measures adult root traits (Trachsel et al., 2010), is time consuming, destructive, and laborious, and limits the number of experiments that can be completed in a season (Trachsel et al., 2010). An alternative to adult root phenotyping would be to study roots at a seedling stage. The relationship between seminal root biomass in hydroponics and root lodging in a field study focusing on root strength and pulling resistance has been explored (Landi, 1998; Landi et al., 2001). Correlations found in hydroponic seedling root traits compared to adult field traits were r=0.44* for shoot weight and adult plant height, and r=0.22* for lateral root length with brace root development In another study, positive but low correlations were found between maize seedling and adult root traits, such as number of seminal roots and weight of seminal roots to root pulling resistance (r=0.07 and r=0.36*, respectively) (Nass and Zuber, 1971). Seedling phenotyping takes less time, is less laborious, and can be repeated many times during the year allowing for quicker turnover of results. Studying seedling root systems also allows to utilize digital image capture systems with more ease. Using digital imaging software to automate phenotypic analysis is an innovative and efficient way of accurately taking

measurements of plant physiological traits (Brewer et al., 2006, Chavarria-Krauser et al., 2008, Wang et al., 2009). With the development of custom root analysis systems, quantitative studies of root architecture are now possible (Le Bot et al., 2009, Zeng et al., 2008). Programs such as of RootReader2D and RootReader3D are examples of imaging software programs developed and made freely available that can be implemented easily without costly equipment (Clark et al., 2013). Expanding the number of seedling root traits and improving respective phenotyping procedures, may increase the chance of capturing strong relationships between different growth stages in maize.

Utilizing roots for crop improvement

Combining both phenotypic information and genomic information is the basis for Genomic Selection or Prediction (GS and GP). The ability to predict high performing genotypes with high accuracy is of major importance in plant breeding. Marker assisted selection (MAS) has become a routine procedure in commercial breeding programs because of an increased gain per unit time by using MAS when compared to only using phenotypic selection (Eathington et al., 2007). GS, a form of MAS, has become a valuable tool in animal breeding, but has yet to be widely implemented in public plant breeding programs. Unlike MAS, where only markers known to be associated with a given trait are used, GS uses all markers simultaneously in order to capture the maximum amount of genetic variation possible to make predictions based on genotype alone. Success of genomic selection procedures depends on saturating the genome with sufficient markers (SNPs) to accurately capture all marker and haplotype effects for prediction of the breeding value of selected genotypes by a three-step process. The first step in conducting genomic prediction is the selection of a training population, one that should be a good representation of the breeding population for accurate estimation of genetic effects (Heffner, et

al., 2009) This population will contain information on the phenotypic traits of interest as well as genome-wide distributed markers. The second step is to establish a statistical model to predict trait performance in a validation (breeding) population based on estimated marker effects alone. These estimated marker effects are determined using the before mentioned prediction model. Accuracy of predicting performance in both plants and animals is affected by many factors. Examples are the extent of LD within both training and validation populations, genetic relationships between the validation and selection population, the genetic architecture of the traits being selected upon, the marker density, the training population size and also the heritability of the traits being studied (Hayes et al., 2009; Luan et al., 2009; Zhong et al., 2009). Finally, the third step is validation of predicted performance for selected genotypes. This information can then be combined with previous training data. Genomic prediction is a tool that could be utilized to predict and select root architectural traits in order to reduce input into resource intensive practices involved in plant root phenotyping.

Optimum root architecture (ideotype) for all environmental conditions has yet to be defined. Lines with better root architecture in relation to root surface area, greater total root length, or various other root traits, might help to improve the ability to take up water and nutrients and consequently, to increase plant growth under nutrient deficiency, in particular for N (Marschener, 1998). A root ideotype for N and water deficient soils was outlined by Lynch (2013) as “steep, cheap, and deep” in accordance to how water and N availability is generally greater in deeper soil strata. Conversely, according to resource allocation theory, plants expending more energy and resources into below ground biomass may have less resources into developing large above ground biomass, so a tradeoff is connected with increasing root biomass

for efficient plant development (Werner, 1981). It is thus critical; to understand what root trait characteristics are optimal for efficient growth and plant productions.

The overall goal of this work was to explore root trait variation between maize inbred lines, identify putative causative loci within already characterized root development genes as well as putatively associated loci throughout the genome involved in root development, and determine the predictability of root traits within a controlled environment model system based on genotypic information alone. The hypothesis of this study is that root traits and root development are quantitatively inherited with many loci throughout the genome associated with root architecture development. Thus, the first objective (Chapter 2) in this project was to determine the impact of known root genes on a range of root phenotypes and identify potential causative loci. We did this by evaluating 74 inbred lines for root traits conducting a candidate gene based association analysis at the seedling stage with plants grown using a paper roll method (Woll et al 2005). The second objective (Chapter 3) was to develop a new tool for root trait extraction from images to enable more high-throughput root phenotyping for large scale mapping studies. The third objective (Chapter 4) was to expand upon objective one and survey for additional genes and QTL beyond known root development genes tested in Chapter 2. To do this we conducted a genome-wide association analysis based on 22 seedling root traits within an inbred line association panel. The final objective (Chapter 5) of this research was to determine whether root traits could be subjected to a genomic prediction approach using the genome-wide association panel of Chapter 4 as a training population, and predict root trait performance in a larger genotyped panel. The goal of this latter approach was to determine whether one could use GP to identify two sets of extreme genotypes using a subset of lines from a larger population, as well as test the predictability of seedling root architecture for TRL as a model for future studies.

Organization of the thesis

This thesis contains three published research articles (Chapters 2-4) and one manuscript in preparation (Chapter 5). The conclusions of all studies are summarized in a final chapter (Chapter 6). As each chapter contains its own introduction, the general introduction was kept brief. Literature for each individual experiment and procedure is introduced and discussed within the respective chapters.

Author Contributions

Chapter 2

BK was first coauthor of study and helped design and carry out all experimental procedures including measurements, data analysis, and primary writer of the manuscript.

AG was first coauthor of study and helped design and carry out all experimental procedures including measurements and data analysis.

JP was coauthor, conducted resequencing of candidate gene *Rth3* for candidate gene association analysis, analyzed data and made major revisions for publication.

JRM helped in design of the experiments and line selection.

FH helped select candidate genes for study and was integral in his expertise in the maize root system.

TL was PI of BK and JP and was involved in design of the experiment and data analysis. TL was also involved in writing the manuscript.

Chapter 3

JP designed all experiments, carried out phenotypic measurements, data analysis and was primary writer of the manuscript.

NL helped analyze images using *ARIA* and wrote much of the scripts for the program.

HSK helped analyze images using *ARIA* and wrote much of the scripts for the program.

BG is PI of NL and HSK and was heavily involved in advice for programing as well as developing parameters for the program.

TL was PI of JP and helped design the experiment as well and discuss results. TL was heavily involved in writing the manuscript.

Chapter 4

JP helped design the experiment, carried out all phenotypic measurements, data analysis, and was the primary writer of the manuscript.

CG helped in selecting lines to be used within mapping population as well as helped in revising written manuscript.

CR was instrumental in supplying genotypic marker data as well as helping with filtering data sets. Also helped with revising the manuscript and gave pointers for GWAS analysis.

BG was lead architect of developing the image analysis software *ARIA*. Also gave feedback on manuscript.

TL was PI of JP and helped design the experiment as well as look at data analysis with JP. TL was heavily involved with writing the manuscript.

Chapter 5

JP designed all experiments, carried out phenotypic measurements, data analysis and was primary writer of the manuscript.

XY helped with data management and filtering of genotypic data as well as advised JP throughout study.

TL was PI of JP and helped design the experiment as well as look at data analysis with JP. TL was heavily involved with writing the manuscript.

References

Aiken, R. M. and A. J. M. Smucker. 1996. ROOT SYSTEM REGULATION OF WHOLE

PLANT GROWTH1. Annual Review of Phytopathology 34: 325-346.

doi:doi:10.1146/annurev.phyto.34.1.325.

Bais, Harsh P., Tiffany L. Weir, Laura G. Perry, Simon Gilroy and Jorge M. Vivanco. 2006.

THE ROLE OF ROOT EXUDATES IN RHIZOSPHERE INTERACTIONS WITH PLANTS

AND OTHER ORGANISMS. Annual Review of Plant Biology 57: 233-266.

Brewer, Marin Talbot, Lixin Lang, Kikuo Fujimura, Nancy Dujmovic, Simon Gray and Esther van der Knaap. 2006. Development of a Controlled Vocabulary and Software Application to Analyze Fruit Shape Variation in Tomato and Other Plant Species. Plant physiology 141: 15-25..

Buckler Iv, Edward S. and Jeffry M. Thornsberry. 2002. Plant molecular diversity and applications to genomics. Current Opinion in Plant Biology 5: 107-111..

Chavarria-Krauser, A., K. A. Nagel, K. Palme, U. Schurr, A. Walter and H. Scharr. 2008. Spatio-temporal quantification of differential growth processes in root growth zones based on a novel combination of image sequence processing and refined concepts describing curvature production. New Phytol 177: 811-821.

Clark, R. T., A. N. Famoso, K. Zhao, J. E. Shaff, E. J. Craft, C. D. Bustamante, et al. 2013. High-throughput two-dimensional root system phenotyping platform facilitates genetic analysis of root growth and development. Plant Cell Environ 36: 454-466.

Drew, M. C. 1975. Nutrient Supply and the Growth of the Seminal Root System in Barley II. Localized, compensatory increases in lateral root growth and rates of nitrate uptake when nitrate supply is restricted to only part of the root system. Journal of experimental botany 26: 79.

Eathington, Sam R., Theodore M. Crosbie, Marlin D. Edwards, Robert S. Reiter and Jason K. Bull. 2007. Molecular Markers in a Commercial Breeding Program. Crop Science 47: S-154.

Esau K. 1965 Plant anatomy. New York Wiley

Feldman, L. 1994. The Maize Root. In: M. Freeling and V. Walbot, editors, The Maize Handbook. Springer New York. p. 29-37.

Hammer, Graeme L., Zhanshan Dong, Greg McLean, Al Doherty, Carlos Messina, Jeff Schussler, et al. 2009. Can Changes in Canopy and/or Root System Architecture Explain Historical Maize Yield Trends in the U.S. Corn Belt? Crop Science 49: 299.

Hayes, B. J., P. J. Bowman, A. C. Chamberlain, K. Verbyla and M. E. Goddard. 2009. Accuracy of genomic breeding values in multi-breed dairy cattle populations. Genet Sel Evol 41: 51.

Heffner, Elliot L., Mark E. Sorrells and Jean-Luc Jannink. 2009. Genomic Selection for Crop Improvement. Crop Science 49: 1.

Hirel, B., J. Le Gouis, B. Ney and A. Gallais. 2007. The challenge of improving nitrogen use efficiency in crop plants: towards a more central role for genetic variability and quantitative genetics within integrated approaches. J Exp Bot 58: 2369-2387.

Hochholdinger, Frank, editor. 2009. Handbook of Maize: Its Biology.

Hoppe, D. C., M. E. McCully and C. L. Wenzel. 1986. The nodal roots of Zea: their development in relation to structural features of the stem. Canadian Journal of Botany 64: 2524-2537.

Landi, P, MM Giuliani, LL Darrah, R Tuberosa, S Conti and MC Sanguineti. 2001. Variability for root and shoot traits in a maize population grown in hydroponics and in the field and their relationships with vertical root pulling resistance. Maydica 46: 177-182.

Landi, P. 1998. Seedling characteristics in hydroponic culture and field performance of maize genotypes with different resistance to root lodging. Maydica 43: 111.

Le Bot, Jacques, Valérie Serra, José Fabre, Xavier Draye, Stéphane Adamowicz and Loïc Pagès. 2009. DART: a software to analyse root system architecture and development from captured images. Plant and Soil 326: 261-273.

Liu, Jianchao, Jiansheng Li, Fanjun Chen, Fusuo Zhang, Tianheng Ren, Zhongjuan Zhuang, et al. 2008. Mapping QTLs for root traits under different nitrate levels at the seedling stage in maize (Zea mays L.). Plant and Soil 305: 253-265.

Luan, T., J. A. Woolliams, S. Lien, M. Kent, M. Svendsen and T. H. Meuwissen. 2009. The accuracy of Genomic Selection in Norwegian red cattle assessed by cross-validation. Genetics 183: 1119-1126.

Lynch, Jonathan. 2013. Steep, cheap and deep: an ideotype to optimize water and N acquisition by maize root systems. Annals of Botany. doi: 10.1093/aob/mcs293

Lynch, Jonathan. 1995. Root Architecture and Plant Productivity. Plant physiology 109: 7-13.

Marschener, H. 1998. Role of root growth, arbuscular mycorrhiza, and root exudates for the efficiency in nutrient acquisition. Field Crops Research 56: 203-207.

McCully, M.E. and M.J. Canny. 1988. Pathways and processes of water and nutrient movement in roots. Plant and Soil 111: 159-170.

Merkangas, Neil Risch and Kathleen. 1996. The Future of Genetic Studies of Complex Human Diseases. Science 273.

Metzker, M. L. 2010. Sequencing technologies - the next generation. Nat Rev Genet 11: 31-46.

Nass, H. G. and M. S. Zuber. 1971. Correlation of Corn (Zea mays L.) Roots Early in Development to Mature Root Development1. Crop Sci. 11: 655-658.

Rafalski, Antoni. 2002. Appications of single nucleotide polymorphisms in crop genetics. Current Opinion in Plant Biology 5: 94-100.

Ribaut, Jean-Marcel, Yvan Fracheboud, Philippe Monneveux, Marianne Banziger, Mateo Vargas and Changjian Jiang. 2007. Quantitative trait loci for yield and correlated traits under high and low soil nitrogen conditions in tropical maize. Molecular Breeding 20: 15-29.

Salvi, R. Tuberosa and S. 2007. From QTLs to Genes controlling Root Traits in Maize. p. 13-22.

T. Presterl, G. Seitz, M. Landbeck, E.M. Thiemt, W. Schmidt, H. H. Geiger. 2003. Improving Nitrogen-Use Efficiency in European Maize: Etimation of Quantitative Genetic Parameters. Crop Science 43: 1259-1265.

Taramino, G., M. Sauer, J. L. Stauffer, Jr., D. Multani, X. Niu, H. Sakai, et al. 2007. The maize (Zea mays L.) RTCS gene encodes a LOB domain protein that is a key regulator of embryonic seminal and post-embryonic shoot-borne root initiation. Plant J 50: 649-659.

Tian, Qiuying, Fanjun Chen, Fusuo Zhang and Guohua Mi. 2006. Genotypic Difference in Nitrogen Acquisition Ability in Maize Plants Is Related to the Coordination of Leaf and Root Growth. Journal of Plant Nutrition 29: 317-330.

Trachsel, Samuel, Shawn M. Kaeppler, Kathleen M. Brown and Jonathan P. Lynch. 2010. Shovelomics: high throughput phenotyping of maize (Zea mays L.) root architecture in the field. Plant and Soil 341: 75-87.

Von Behrens, I., M. Komatsu, Y. Zhang, K. W. Berendzen, X. Niu, H. Sakai, et al. 2011. Rootless with undetectable meristem 1 encodes a monocot-specific AUX/IAA protein that controls embryonic seminal and post-embryonic lateral root initiation in maize. Plant J 66: 341-353.

W.R. Jordan, W.A. Dugas, P.J. Shouse. 1983. Strategies for Crop Improvement for Drought-Prone Regions. Agriculture Water Management 7: 281-299.

Wang, Liya, Ioan Vlad Uilecan, Amir H. Assadi, Christine A. Kozmik and Edgar P. Spalding. 2009. HYPOTrace: Image Analysis Software for Measuring Hypocotyl Growth and Shape Demonstrated on Arabidopsis Seedlings Undergoing Photomorphogenesis. Plant physiology 149: 1632-1637.

Watt, Michelle, Wendy K. Silk and John B. Passioura. 2006. Rates of Root and Organism Growth, Soil Conditions, and Temporal and Spatial Development of the Rhizosphere. Annals of botany 97: 839-855.

Werner, J. D. Soule and P.A. 1981. Patterns of Resource Allocation in Plants, with Special Reference to Potentilla recta L. Torrey Botanical Society 109: 311-319.

Yan, Jianbing, Marilyn Warburton and Jonathan Crouch. 2011. Association Mapping for Enhancing Maize (L.) Genetic Improvement. Crop Science 51: 433.

Zeng, G., S. T. Birchfield and C. E. Wells. 2008. Automatic discrimination of fine roots in minirhizotron images. New Phytol 177: 549-557.

Zhong, S., J. C. Dekkers, R. L. Fernando and J. L. Jannink. 2009. Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a Barley case study. Genetics 182: 355-364.

Zhu, J., S. M. Mickelson, S. M. Kaeppler and J. P. Lynch. 2006. Detection of quantitative trait loci for seminal root traits in maize (Zea mays L.) seedlings grown under differential phosphorus levels. TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik 113: 1-10.

CHAPTER TWO

ASSOCIATION ANALYSIS OF SINGLE NUCLEOTIDE POLYMORPHISMS IN CANDIDATE GENES WITH ROOT TRAITS IN MAIZE (*ZEA MAYS* L.) SEEDLINGS

Bharath Kumar[†1], Adel H. Abdel-Ghani[†2*], Jordon Pace[1], Jenaro Reyes-Matamors[3], Frank Hochholdinger[4] and Thomas Lübberstedt[1]

Paper published in Plant Science Journal. Abstract, structure, and references are all formatted according to journal standards.

ABSTRACT

Several genes involved in maize root development have been isolated. Identification of SNPs associated with root traits would enable the selection of maize lines with better root architecture that might help to improve N uptake, and consequently plant growth particularly under N deficient conditions. In the present study, an association study (AS) panel consisting of 74 maize inbred lines was screened for seedling root traits in 6, 10, and 14-day-old seedlings. Allele re-sequencing of candidate root genes *Rtcl*, *Rth3*, *Rum1*, and *Rul1* was also carried out in the same AS panel lines. All four candidate genes displayed different levels of nucleotide diversity, haplotype diversity and linkage disequilibrium. Gene based association analyses were carried out between individual polymorphisms in candidate genes, and root traits measured in 6, 10, and 14-day-old maize seedlings. Association analyses revealed several polymorphisms within the *Rtcl*, *Rth3*, *Rum1*, and *Rul1* genes associated with seedling root traits. Several nucleotide polymorphisms in *Rtcl*, *Rth3*, *Rum1*, and *Rul1* were significantly ($P<0.05$) associated with seedling root traits in maize suggesting that all four tested genes are involved in the maize root

development. Thus considerable allelic variation present in these root genes can be exploited for improving maize root characteristics.

1. Introduction

The plant root system serves primarily to anchor plants in the soil, and to take up water and minerals. Roots are less visible than aboveground plant parts such as flowers, stems, and leaves. Therefore, root characteristics are seldomly considered as selection criteria [1], but they are no less important to the plant. The root system is affected by environmental conditions, management practices, and to a greater extent genotype dependent. While plants respond to limiting soil nutrients and water stress by increasing the amount of root biomass allocated to roots, and consequently increasing root to shoot biomass ratio [2-7], the acquisition of soil nutrients and available soil moisture by plants is more dependent upon root length and/or root surface area than total root biomass [8-9]. Genetic variation for root morphology in maize does exist, and has long been proposed for improvement of nutrient and water-use efficiency in maize [7, 10-14].

Root architecture traits can be determined using different methods including vertical root pulling force (RPF) and hydroponic characterization [15-18]. Field methods are frequently technically demanding and costly. Due to the difficulty in obtaining reliable root trait data from the field, there are very few reports on morphological characterization of maize roots in the field. Using paper rolls as a hydroponic method to study root architecture has several advantages in comparison with RPF and other field techniques [7, 14, 18-19]. These include: (i) the ease to score root traits as compared with vertical RPF, (ii) controlled environmental conditions, thus increasing repeatability of measurements, (iii) screening large numbers of lines in small space within a short period of time with an easy access to roots, and (iv) precise control of the

concentration of mineral nutrients and water soluble compounds. However, the main disadvantages are the artificial screening conditions which might not properly represent field conditions.

Maize varieties with high yield potential are expected to have favorable root architecture, which can effectively supply water and nutrients, leading to increased grain yield [7, 15-17]. This is particularly important in case of limited water or nutrient availability, such as under drought conditions.

The maize root system consists of different root types that are formed during different stages of plant development. The root system in maize can be divided into embryonic and post-embryonic roots [20]. The embryonic root system is composed of a single primary root and a variable number of seminal roots, while post-embryonic roots are shoot-borne roots including crown and brace roots. Shoot-borne roots formed at consecutive underground nodes are called crown roots, while the respective roots formed at consecutive above-ground nodes of the shoot are called brace roots. Lateral roots which emerge from all major root-types also belong to the post-embryonic root system. Mutants affected in various aspects of root formation have been identified in maize including *rtcs*, *rth1*, *rth3,* and *rum1* [19, 21-23]. *Rtcl* (*Rtcs-like*) is regarded as a paralog of *Rtcs* [22], and *Rul1* (*Rum1-like*) as a paralog of *Rum1*. That the primary root and its lateral roots alone are sufficient to form a fertile mature plant was demonstrated by the monogenic recessive mutant *rtcs*, which forms only a primary root and its lateral roots but no seminal or shoot-borne roots [24]. The mutant *rum1* is affected in lateral root formation, while the mutants *rth1* [21] and *rth*3 [25] display reduced root hair elongation. *Rtcs* encodes a 244 amino acid (aa) Lateral Organ Boundaries (LOB) domain protein located on chromosome 1S. During evolution, *Rtcs* was duplicated. The *Rtcl* gene, which maps on chromosome 9, displays

72% aa sequence similarity with *Rtcs*. The *Rtcs* and *Rtcl* gene promoters share auxin responsive elements, and they are preferentially expressed in roots [22]. The *Rth1* gene encodes a SEC3 homologue [21]. In yeast (*Saccharomyces cerevisiae*) and mammals, *sec3* is part of the exocyst complex, which ropes together exocytotic vesicles prior to their fusion. The *Rth3* gene belongs to the COBRA-like gene family [25]. Members of this plant-specific glycosylphosphatidylinositol anchored protein coding gene family are involved in cell expansion and cell wall biosynthesis [26]. The *Rum1* gene located on chromosome 3 encodes a polypeptide of 269 aa which is a monocot specific AUX/IAA protein [23]. *Rul1* is a closely related Aux/IAA protein coding gene, and is localized on chromosome 8. *Rul1* encodes a polypeptide of 273 aa that displays 92% aa identity with *Rum1*.

Recombinant inbred lines have been used for mapping quantitative trait loci (QTL) to 10-30 cM regions [27-28]. However, QTL mapping is limited by, (i) the expense of generating such lines, (ii) their limited diversity, (iii) their separation from established processes in maize breeding, and (iv) the low number of informative recombinations [29]. In contrast, association mapping studies which are based on linkage disequilibrium (LD) allow identification of actual genes underlying these QTLs [30]. The success of gene based association studies depends on the candidate gene(s) chosen for a particular phenotypic trait. The first candidate gene-based association mapping study in maize associated individual *dwarf8* polymorphisms with flowering time [30], which has been followed by numerous subsequent studies in maize [31] and other cereal crops [32]. Gene-based association studies ultimately lead to the identification of quantitative trait polymorphisms (QTPs) with causal genetic effects on agronomic traits, which can be converted into functional markers [33]. Breeding for a vigorous root system in maize may involve identification of superior alleles of candidate genes that affect nutrient and water use efficiency. Respective

candidate gene-based studies enabled identification of alleles affecting various relevant quantitative agronomic traits in maize [30, 34-38].

So far, no information is available on the genetic diversity of genes affecting root development in maize. Therefore, the objectives of this study were to: (i) examine the nucleotide and haplotype diversity for *Rtcl*, *Rth3*, *Rum1*, and *Rul1* in a panel of 74 maize inbreds, (ii) estimate phenotypic means for root traits of lines included in the individual haplotypes, and (iii) to identify polymorphisms in candidate genes associated with root development.

2. Materials and methods

2.1. Plant materials

Allele re-sequencing of candidate root genes was carried out in 44 expired PVP lines, and 30 public inbred lines such as Nested Association Mapping (NAM) founder lines, 2009 released Germplasm Enhancement of Maize (GEM) lines and lines used in a maize diversity study (Appendix 1). The rationale for using expired PVP lines is to capture substantial genetic variation present in current elite germplasm. Other public inbred lines were chosen to enable detection of the majority of SNP and INDEL polymorphisms in the candidate genes studied, as a prerequisite to develop multiplexed SNP assays to be used for screening large numbers of genotypes at low costs in large-scale association studies. Seed was obtained from different seed resource centers such as North Central Regional Plant Introduction Station in Ames, IA (NCRPIS), and Maize Genetics Cooperation (Champaign, IL). All maize lines were selfed at the Agronomy farm, Iowa State University in summer 2009 to produce seed of equal origin and quality for this study.

2.2. Experimental design and phenotyping

Seedling root characteristics in maize lines were studied using a paper roll test described by Woll et al. [19]. Seeds were first surface sterilized with Clorox® solution (6% sodium hypochlorite) for 15 minutes. After surface sterilization, seeds were washed three times with sterile water. Surface sterilized seeds were then placed on a brown germination paper (Anchor Paper, St. Paul, MN) pre-moisturized with fungicide solution Captan® (2.5g/l), and afterwards rolled up vertically. Rolled germination papers were kept in 2 l glass beakers containing autoclaved deionised water. Experiments were carried out in growth chambers under a photoperiod of 16/8 h at 25/22 $^{\circ}$C (light/darkness) with photosynthetically active radiation of 200 µmol photons m^{-2} s^{-1}. The relative humidity in growth chambers was maintained at 65%, and lines were grown in a randomized complete block design with two replications. Each paper roll containing three seedlings was considered an experimental unit. 74 maize lines with different genetic background and origins were evaluated at three growth stages independently (6, 10, and 14 days after sowing). Each experiment was repeated twice. At the end of each growth stage (6, 10, and 14 days), root characteristics were evaluated. Seedlings were separated into root system and shoots at the crown root region. The root system was further separated into primary root, seminal, and crown roots, and respective root lengths were recorded. To measure lateral roots, the primary root was scanned, and the image was analyzed using WinRhizo Pro 2009 software (Regent Instruments, QC, Quebec, Canada). Total root length (TRL) was estimated by summing the lengths of primary root, crown, seminal, and lateral roots for each seedling. Roots were dried at 70 °C to a constant weight, and root dry weight (RDW) was recorded subsequently.

2.3. DNA extraction, amplification, and sequencing

Four candidate genes were chosen based on their role in root development to identify SNPs for association study analyses. SNPs from these candidate genes were tested for possible associations with TRL and RDW. Candidate genes chosen for our association study were: *Rtcs*, *Rtcn*, *Rtcl*, *Rth3*, *Rum1,* and *Rul1*. Gene specific primers were designed to amplify the entire sequence of *Rtcl,* and parts of *Rth3*, *Rum1,* and *Rul1* genes using the software program Primer 3.0 (http://frodo.wi.mit.edu/primer3/) (Table 3). In case of *Rtcs* and *Rtcn*, even after several attempts, amplicons from all 74 lines were not obtained. This might be due to the extensive nucleotide diversity at these candidate genes which prevents the binding of designed primers. Polymerase chain reaction (PCR) was performed using the designed primers for each gene separately in 50 μl volumes under the following conditions: 50 ng template DNA, 250 nM of each primer, 250 nM dNTPs, 2 U Taq polymerase and 250 μM $MgCl_2$. Reactions were performed for each primer pair using the following PCR program in a thermocycler (MJ research, California): an initial 94 °C denaturation step for 2 min followed by 35 cycles of 94 °C for 30 sec (denaturation step), 57.5 °C for 30 sec (annealing step), and 72 °C for 90 sec (elongation step). The final extension step was followed by 72 °C incubation for 10 min. Amplified DNA fragments were resolved by gel electrophoresis (Biorad, California) using 1% agarose gels in Tris-EDTA (TE) buffer. Agarose gels were stained with 0.5 µg of ethidium bromide per ml. The running time was 90 min at 120 mV. Finally, gels were visualized and photographed by a UV illuminator system (Alphainnotech, California). For each gel, the first lane was specified for a 100 bp DNA ladder (Promega, Wisconsin), the second lane and the third lane were specified for positive and negative controls. Amplified fragments of *Rth3, Rum1,* and *Rul1* genes were obtained for all 74 inbred lines in the AS panel, whereas for the *Rtcl* gene,

amplicons were obtained from 69 lines. For sequencing, 10 µl of the amplified fragments were first purified by using 2 units of shrimp alkaline phosphatase and 2 units of exonuclease I at 37 °C for 1 h, followed by 72 °C for 15 min to deactivate the enzymes. Amplified gene products were then labeled for sequencing using the ABI Prism® BigDye® Terminator v3.1 cycle sequencing kit (Applied Biosystems, California). Labeling reactions were performed in 10 µl reaction volume containing 1 µl of PCR product, 1 µl of BigDye Terminators, 0.26 µl of 50 mM original PCR primers (either forward or reverse), 1.75 µl of 5× sequencing buffer and 5.99 µl deionized distilled water. The thermocycler cycle sequencing reaction was performed using the following cycling parameters: 96 °C for 2 min, 25 cycles of 96 °C for 30 sec, 50 °C for 1 min, and 60 °C for 4 min, finally cooled to 4 °C. Precipitated DNA was purified with 70% ethanol and dried thoroughly before re-suspending in ABI Hi-Dye formamide for sequencing on a sequencer (Applied Biosystems 3730 DNA Analyzer with a 96-capillary array). Sequencing was performed for each amplified fragment using forward and reverse primers separately with two replicates. Based on primers designed, expected sizes of PCR products were obtained for all tested genes. Sequences were aligned using Sequencher program 4.1 (Gene Codes Corporation, Michigan). In order to maximize read lengths and to obtain a sequencing quality >98%, two replicates of forward and reverse reads for each amplified fragment were aligned to get consensus sequences of amplified gene fragments from AS panel lines.

2.4. Phenotypic data analyses

The following linear mixed model was used to estimate variance components: $y_{ijk} = \mu + E_i + B_{j(i)} + G_k + EG_{ik} + BG_{j(i)}k$, where y_{ijk} represents the observation from the ijk^{th} experimental unit, μ is the overall mean, E_i is the effect of i^{th} independent experiment, $B_{j(i)}$ is the effect of j^{th} block nested in i^{th} experiment, G_k is the effect of k^{th} line, EG_{ik} is the interaction effect of the i^{th}

experiment with k^{th} genotype, $BG_{j(i)k}$ is the interaction effect of j^{th} block nested in i^{th} experiment with k^{th} genotype. Heritability (h^2) on an entry mean basis was estimated as the ratio of genotypic to phenotypic variance according to Hallauer and Miranda [39]. Furthermore, experiment was considered as a fixed factor, whereas blocks and genotypes were regarded as random factors. Best linear unbiased estimates (BLUEs) were determined for maize lines for each trait. SAS 9.1 (SAS Institute, 1996) software packages were used for all calculations.

2.5. Analysis of sequence data

Respective gene sequences amplified from the association panel lines were analyzed using the software package DnaSp [40]. Haplotype diversity among candidate genes was analyzed based on the SNPs in the amplified fragment sequences from AS panel lines. Levels of nucleotide diversity in different parts of amplified fragments of the gene were estimated as π, the average number of nucleotide differences per site between two sequences [41]. A different estimator of nucleotide diversity θ, the neutral mutation parameter was calculated based on number of segregating sites [42] with a common expected value $\theta = 4N_e\mu$, where N_e equals the effective population size and μ the mutation rate per generation and site. Haplotype diversity (*Hd*) was estimated as the probability that two randomly chosen haplotypes from a given population were different [43]. Neutrality of mutations was checked using Tajima's D statistics [44, 45]. These statistics are based on the different comparisons of $\theta = 4N_e\mu$, where N_e equals the effective population size and μ the mutation rate [42]. Tajima's D statistics results from the comparison of θ based on number of pair-wise differences and the number of segregating sites between sequences in the sample.

2.6. Population structure and association analysis

All 74 lines in the AS panel were genotyped with 101 SNP markers distributed evenly across 10 maize chromosomes [46] to assess and control the effect of population structure. The software package *Structure 2.0* [47] was used to estimate population structure (Q) within the AS panel using SNP data. In *Structure 2.0*, a burn-in length of 50.000 followed by 50.000 iterations for each of the clusters (K) varying from 1 to 20 (each K was run 20 times) were used to produce a Q matrix estimating membership coefficients for each line in each subpopulation. The Admixture model was applied with independent allele frequencies. An ad hoc (ΔK) statistic [48] based on the second order rate change of P (X|K) was used to identify the most probable value of K.

Loiselle kinship coefficients between lines (a K matrix) were estimated by the TASSEL program [49] based on the 101 SNP markers. Both Q matrix and a K matrix were used in the association analysis to control the spurious associations due to population structure and relatedness, respectively [50]. Association analysis between SNPs and root traits was carried out using a mixed linear model (MLM) implemented in the program *TASSEL 2.10* software [49]. The MLM accounts for overall population structure (Q) and for finer scale relative kinship (K). The statistical model used in mixed linear model (Q+K) can be described in Henderson's notation [51] as: $y= X\beta+Zu+e$, where y is the vector of observations; β is an unknown vector containing fixed effects including genetic marker and population structure (Q); u is an unknown vector of random additive genetic effects from multiple background QTL for individuals or lines; X and Z are the known design matrices; and e is the unobserved vector of random residuals.

TRL and RDW were measured in 6 (sTRL, sRDW), 10 (tTRL, tRDW), and 14 (fTRL, fRDW) day old seedlings, and used as root traits in our association study. False discovery rate was set at

0.05 [52] to control for multiple testing of SNP markers. Motifs in the *Rtcl*, *Rth3*, *Rum1,* and *Rul1* genes were searched using a PLACE (Plant cis-acting regulatory DNA elements) database [53] to determine, if any of the significantly associated SNPs might alter motif sequences in the candidate genes.

3. Results

3.1. Phenotypic variation

Complete statistical analysis of root traits measured in this study has been reported elsewhere [14]. In their principal component analysis, total root length (TRL) and root dry weight (RDW) explained most of the phenotypic variation. Moreover, both TRL and RDW were significantly and positively correlated with all other root-related traits. We, therefore, focused on TRL and RDW for association mapping in this study.

Frequency distribution of TRL and RDW measured in 6, 10, and 14-day-old seedling of 74 AS panel maize inbred lines are presented in supplementary figures 1 and 2. TRL ranged from 8.1-72.6 cm, 39.2-216.3 cm and 78.6-362.0 cm in 6, 10, and 14-day-old seedlings, respectively. RDW varied from 5.5-29.8 mg, 10.1-49 mg, and 14.9-82.0 mg in 6, 10, and 14-day-old seedlings, respectively. Both TRL and RDW had the highest co-efficient of variation (CV) in 6-day-old seedlings.

3.2. Sequence alignment and haplotypes

The *Rtcl* sequence alignment of 69 maize lines spanned 828 bp with no alignment gaps, such as indel polymorphisms. The 828 bp amplified fragment included two exons, i.e., exon 1 (420 bp), and exon 2 (279 bp), respectively, separated by an intron (129 bp). In exons 1 and 2, 16 and 22 SNPs were identified, respectively, whereas 7 SNPs were identified in the intron region. Out of the 38 SNPs in the exon regions, 32 altered the amino-acid sequences; the other 8 were

synonymous mutations (Table 1). In case of *Rth3*, 714 bp of the open reading frame (ORF) region of the gene was amplified from all 74 lines in the AS panel. Sequence alignment of *Rth3* amplicons resulted in the identification of 15 SNPs with no indel polymorphisms. Out of 15 SNPs, 12 were synonymous mutations, and only 3 SNPs altered the amino-acid sequence (Table 2). Intron 4 and exon 5 were partially amplified for *Rum1*: 225 out of 461 bp in intron 4 and 207 out of 315 bp in exon 5. Sequence alignment of *Rum1* amplicons from all 74 lines in the AS panel resulted in the identification of 12 SNPs. Out of 12 SNPs, 9 SNPs were from intron 4 and remaining 3 SNPs were from exon 5 region. Out of 3 SNPs identified in the exon region of *Rum1* gene, two were synonymous mutations and the remaining one SNP altered the amino acid sequence (Table 3). For *Rul1*, intron 5 and exon 6 were partially amplified from all 74 lines of the AS panel. Sequence alignment of 411 bp from *Rul1* consisting of 84 bp of intron 5 and 327 bp of exon 6 resulted in the identification of six SNPs, including five in exon 6. Four of those exon SNPs altered the amino-acid sequence (Table 4).

The number of haplotypes for the four genes ranged from 7 for *Rul1*, 9 for *Rth3*, 16 for *Rtcl*, to 22 for *Rum1* (Tables 1-4). The range of haplotype means for TRL and RDW traits measured in 6, 10 and 14-day-old seedlings was larger for *Rum1* gene compared to other three genes (Table 1-4).

3.3. Nucleotide diversity in four genes

Nucleotide diversity (π) was determined for *Rtcl, Rth3, Rum1,* and *Rul1* coding and non-coding regions using the SNPs identified in respective amplicons from AS panel lines (Table 5). Overall, nucleotide diversity was π=0.021 in the entire region of *Rtcl*. Within *Rtcl*, nucleotide diversity was almost the same in both intron (π=0.022) and exon (π=0.021) regions. In *Rth3* which lacks an intron region, nucleotide diversity was higher for synonymous (π=0.026) than for

non-synonymous mutations (π=0.0002). For the *Rum1*, nucleotide diversity was higher in the non-coding region (π=0.017) than in the coding region (π=0.005), and for *Rul1*, there was not much difference in the nucleotide diversity between non-coding (π=0.007) and coding region (π=0.004). When the entire amplified region was considered, nucleotide diversity was lower in *Rum1* (π=0.011), *Rth3* (π=0.007), and *Rul1* (π=0.005) compared to *Rtcl* (π=0.021). The nucleotide diversity based on θ, the neutral mutation parameter, was also calculated for all four amplicons in a sliding window of 100 bp using a step size of 10 bp (Fig. 1). Based on θ, within *Rtcl*, average nucleotide diversity was same in both intron and exon region. In case of the *Rum1* gene, nucleotide diversity seems to be higher in the intron region compared to the exon, but it was the same in both the exon and the intron region in the amplified region of the *Rul1* gene. Haplotype diversity (*Hd*) ranged from 0.873 in *Rtcl* to 0.624 in *Rul1*.

Tajima's *D* was positive and significant when considering the entire *Rtcl* region as well as both coding and non-coding regions. Conversely, in case of *Rth3*, *Rum1* and *Rul1*, Tajima's *D* was non-significant in all regions except in the *Rul1* non-coding region (Table 5). Complete analysis of LD decay in AS panel lines across *Rtcl*, *Rth3*, *Rum1* and *Rul1* genes has been reported elsewhere [54]. LD between all pairs of polymorphic sites from the sequenced region of the *Rtcl*, *Rum1*, and *Rul1* genes decayed very rapidly ($r^2 < 0.2$), whereas LD persisted ($r^2 > 0.25$) over the length of the sequenced region in the *Rth3* gene.

3.4. Population structure and marker-trait associations

Based on the ad-hoc statistic values in *Structure 2.0*, lines in the AS panel were grouped into two sub-populations (K=2), which agrees with their pedigree and breeding history [54].

3.4.1. Rtcl

21 SNPs were significantly (P=0.05) associated with sTRL, and 16 SNPs were associated with sRDW (Table 6), with 14 SNPs associated with both sTRL and sRDW. Five of those SNPs were located in exon 1, four in the intron, and five in exon 2. Four SNPs in exon 1 and four in exon 2 caused non-synonymous changes in the protein sequence (Table 6), while the remaining two SNPs in the exon regions caused synonymous changes. In case of tTRL and tRDW, five and two SNPs were associated, respectively. SNPs at the sites 413, 473, 531, 547, and 554 were significantly associated with both sTRL and tTRL. Similarly, SNPs at sites 320 and 374 were significantly associated with both sRDW and tRDW. Out of these SNPs, SNPs at sites 320, 374, 413, and 554 caused non-synonymous changes in the amino-acid sequence. For fourteen-day-old maize seedlings, SNPs at sites 510 and 554 were associated with fTRL only. Moreover, the SNP at site 510 was associated with both sTRL and fTRL, whereas the SNP at 530 was associated with sTRL, tTRL and fTRL.

Using B73 as reference sequence, seven SNPs (290, 317, 320, 468, 510, 597, and 799) were significantly associated either with TRL and RDW traits affected putative functional sequence motifs in the *Rtcl* gene. These motifs are the signatures of the binding sites of several regulatory proteins (Supp. Table 1). Moreover, non-synonymous SNPs at 290, 317, and 320 affected the LOB domain amino acids in the *RTCL* gene (data not shown).

3.4.2. Rth3

13 polymorphisms in the *Rth3* exon region were associated with sTRL, whereas the SNP at 621 was the only polymorphism associated with sRDW (Table 7). Of these associated SNPs, a synonymous polymorphism at the site 621 was associated with both sTRL and sRDW. In case of fTRL and fRDW, seven and eight SNPs were associated, respectively. The synonymous SNPs at sites 180, 234, 438, 465, 492, 519, and 600 were significantly associated with both fTRL and

fRDW. Moreover, these SNPs were also associated with sTRL. No SNP in *Rth3* was associated with tTRL. Four SNPs (389, 399, 436 and 600) significantly associated with TRL and RDW affected the binding sites for regulatory factors in the *Rth3* (Supp. Table 1). Since these SNPs were synonymous, they did not affect the COBRA domain in the *Rth3* gene.

3.4.3. Rum1

One and two SNPs in *Rth3* were associated with sTRL and sRDW, respectively. The SNP at site 303 in the intron 4 region was associated with both sTRL and sRDW. In case of tTRL and tRDW, the SNPs at sites 63 and 251 were associated with both traits. Moreover, these SNPs were also associated with fTRL and fRDW. SNPs at sites 118 and 302 were associated with sTRL and sRDW and also with fTRL and fRDW. The SNP at site 118 in the exon 5 region causes a non-synonymous change in the amino-acid sequence and also affects a binding site transcription factors in the *Rum1* gene (Supp. Table 1).

3.4.4. Rul1

SNPs at sites 311, 336, and 389 in the exon 6 region of *Rul1* were significantly associated with tRDW. The SNPs at sites 336 and 389 caused non-synonymous changes in the amino-acid sequence. A synonymous SNP at site 7 in the intron 5 region of *Rul1* was associated with sRDW. No SNP from the amplified *Rul1* gene region was associated with either fTRL or fRDW.

4. Discussion

4.1. High levels of phenotypic, nucleotide, and haplotype diversity

We observed substantial quantitative variation for root traits TRL and RDW in 6, 10, and 14-day-old seedlings indicating a considerable amount of morphological differences among 74 maize inbred lines in the AS panel (Supp. Figs. 1 and 2). We identified maize lines with both

under and well–developed root systems, which are attractive for identifying genomic regions controlling root traits [14].

In the present study, 2386 bp across four candidate genes involved in root development were amplified from the AS panel lines, resulting in 78 SNPs, and an average SNP frequency of 1 SNP/31bp (Table 5). Substantial differences in nucleotide diversity were observed between the four candidate genes (Table 5). Nucleotide diversity was highest in the *RTCL* gene (π=0.021), and lowest in *Rth3* (π=0.007) and *Rul1* (π=0.005) gene. The nucleotide diversity observed in the candidate genes is comparable to previous studies in maize inbreds for *Sclce2* (π=0.0112; Li et al., 2011), *4CL2* (π=0.0102; [55]), *COMT* (π=0.008; [56]) and *PAL* (π=0.004; [35]). In other studies involving maize landraces, nucleotide diversities ranged from π = 0.001 – 0.0133 with an average value of π=0.004, and a SNP frequency of one SNP per 62 bp. When coding and non-coding regions were compared in candidate genes used in these studies, nucleotide diversity varied across the genes. In case of *Rtcl* and *Rul1*, both intron and exon regions had the same nucleotide diversity, whereas nucleotide diversity was higher in the intron region of *Rum1* gene. This distribution of nucleotide diversity across intron and exon regions has also been found in other studies [55, 57]. All four candidate gene *Rtcl*, *Rth3*, *Rum1,* and *Rul1* showed positive Tajima's D values (Table 5). This indicates balancing selection with an excess of alleles with intermediate frequencies and a scarcity of rare alleles. Considerable haplotype diversity was found for *Rtcl*, *Rth3*, *Rum1,* and *Rul1* (Table 5).

4.2. Polymorphisms associated with root traits

Several studies have shown the quantitative and qualitative importance of root traits in taking up nitrogen (N) from N-depleted soils [58-60]. Identification of the genetic regions associated with root traits would help not only to develop maize lines with a favorable root system, but also to

understand the relationship between plant growth, plant productivity and root architecture. In our previous study, we identified significant positive correlations between seed root traits such as SRL and RDW with grain yield under two N levels [7]. Here, we used association mapping to dissect the role of SNPs in *Rtcl*, *Rth3*, *Rum1,* and *Rul1* for maize root development.

Taramino et al. [22] isolated the first root gene in maize (*Rtcs)* involved in seminal and crown root formation by map-based cloning. *Rtcl*, a paralog of *Rtcs* was used in our association mapping study. The role of *Rtcl* in maize needs yet to be determined. In our association study, *Rtcl* was found to be associated with root development in 6, 10, and 14-day-old seedlings (Table 6). Several synonymous and non-synonymous SNPs in the *Rtcl* gene region were significantly associated with TRL and RDW. This suggests a potential role of *Rtcl* gene in maize root development. This likely role of the *Rtcl* gene in maize root development might be due to the sequence similarity it shares with its paralogous *Rtcs* gene, which has been demonstrated to be involved in root development. The paralogous *Rtcl* gene shares 72% sequence similarity at the protein level with *Rtcs* gene, contains a LOB protein domain, which was found in genes involved in root development [61], and both *Rtcs* and *Rtcl* gene promoters share auxin responsive elements that are preferentially expressed in roots [22]. It has also been shown that maize mutants with impaired LOB domain have reduced crown and seminal roots [62-64].

SNPs in *Rth3* were significantly associated with TRL and RDW in 6 and 14–day-old seedlings. Even though root hair elongation was not measured in this association study, our study suggests that *Rth3* affects other root characteristics in maize. Our findings are consistent with findings of Hochholdinger et al. [25], showing significant yield losses of the *rth3* mutant in replicated field trials. *Rth3* belongs to COBRA – like gene family specifically involved in cell expansion and cell wall biosynthesis [25-26]. The *rth3* mutant has been shown to affect root hair elongation and

grain yield [25]. By this association mapping study we found that *Rth3* affects both TRL and RDW in maize seedlings. The significant association between SNPs in the *Rth3* gene with root length and biomass might be due to the role of root hairs in water and nutrient uptake. Previous studies have shown that plants lacking efficient uptake of water and nutrient have poor root characteristics [59, 65].

Von Behrens et al. [23] isolated the *Rum1* gene that is auxin/indole acetic acid (IAA) inducible and encodes protein containing four conserved domains, and a bipartite nuclear localization sequence. The protein encoded by *Rum1* is involved in the formation of embryonic seminal root and post-embryonic lateral roots. *Rul1* is regarded as paralog of the *Rum1* gene, since it shares 92% sequence identity at the amino acid level and is located in a duplicated region of the maize genome. The role of *Rul1* gene in maize root formation is still unknown. In our association mapping study, *Rum1* was associated with TRL and RDW in 6, 10, and 14-day-old seedlings, thus confirming the role of *Rum1* in maize root development. Moreover, SNPs in the *Rul1* gene were associated with RDW in 10 and 14-day-old seedlings. This suggests a role of *Rul1* in root development, which has so far only been shown to be a paralog of *Rum1* [23].

4.3. Molecular physiological basis of SNP– trait associations

Previous studies have shown the potential role of *Rth3* and *Rum1* genes in maize root development. Any impaired expression of these genes leads to defective root development. From our gene based association study, we not only confirmed the role of *Rth3* and *Rum1* genes in maize root development, but we also found that the two paralogous genes *Rtcl* and *Rul1* are involved in the maize root formation. Thus, it is conceivable that polymorphisms in *Rtcl*, *Rth3*, *Rum1,* and *Rul1* affect maize root formation.

In the *Rtcl* gene, 13 non-synonymous and 4 synonymous SNPs were associated with TRL and RDW. Out of these associated SNPs, seven affected putative functional sequence motifs, mostly transcription factor binding sites. Moreover, out of these seven SNPs, three SNPs at sites 290, 317, and 320 also affected the LOB domain in the *Rtcl* gene. These SNPs seem to be critical not only for the formation of a proper LOB domain, which is required for root formation, but also for regulation of the *Rtcl* gene by affecting transcription factor binding sites. Similar results wherein the SNPs associated with traits affect transcription factor binding sites in the gene have been reported elsewhere [36, 43]. In our previous association mapping study involving SNPs from the *Rtcl* gene and seedling root traits measured under contrasting nitrogen levels, these three SNPs were consistently associated with seedling root traits. The SNP at site 317 in *Rtcl* gene was associated with both RDW and TRL under high and low N conditions, whereas the SNP at site 320 was associated with RDW under both N conditions. In case of the SNP at site 290, associations were observed with RDW and TRL under high N. These consistent associations suggest the potential role of these SNPs in the *Rtcl* gene in maize root development. LD is very low between SNPs at sites 290-320 (r^2 = 0.0255) and 317-320 (r^2 = 0.0903), whereas it was moderate between 290-317 (r^2 = 0.2827). Low to moderate LD between these significant SNPs suggests that these individual SNPs are putative causative polymorphisms, and can be of potential use in deriving markers to select root traits. For the *Rth3* gene, 13 polymorphisms were found to be significantly associated with TRL and RDW. Of these 13 SNPs, four SNPs (site 393, 399, 438, and 600) in the exon region significantly affected the binding sites for regulatory factors in *Rth3*, but none of these SNPs affected the COBRA domain within the gene, as they were synonymous mutations. When the LD was estimated between these four SNPs, low LD was detected between SNPs at the sites 393-438 (r^2 =0.0445), 393-600 (r^2 =0.0323), 399-438

(r^2=0.0445) and 399-600 (r^2 =0.0323). However, there was a high LD between sites at positions 393-399 and 438-600. This suggests, that individual SNPs at 393 (or 399), 438 (or 600) are putative causative polymorphisms, and can potentially be used to derive markers to select root traits. In our previous association mapping study involving SNPs in the *Rth3* gene and grain yield, the SNP at site 600 was associated with grain yield under high N suggesting that this SNP might potentially be used along with other SNPs to select for grain yield. In case of the *Rth3* gene, full-length re-sequencing of this candidate gene would greatly increase the number of unlinked polymorphisms to be tested for associations due to the extent of LD over a long distance.

In our previous association mapping study, non-synonymous SNPs in *Rum1* and *Rul1* gene (site 118 in *Rum1*, 336 and 389 in *Rul1*) were associated with seedling root traits under HN and LN conditions. In the present study, these SNPs were also associated with TRL and RDW. A non-synonymous SNP at site 118 in *Rum1* gene associated with RDW in 6 and 10-day old seedling also affected the putative functional sequence motifs which are the signatures of the transcription factor binding sites in the gene. LD is high between the sites 336-389 in *Rul1* gene, so these polymorphisms are putatively causative with the root trait. Taken together, the SNP at site 118 in the *Rum1* gene, and either SNP at sites 336 and 389 in the *Rul1* gene can potentially be applied in breeding programs to improve root traits.

In the present study, genes and their paralogues have been tested for association with roots traits. From our results, it seems that *Rtcl*, *Rth3*, *Rum1,* and *Rul1* can be considered as candidate genes to develop functional markers for root traits especially the significant SNPs in these genes with large effect on the trait (Supp. Table 2).

Functional markers are DNA markers derived from polymorphic sites within genes, causally involved in phenotypic trait variation [33]. All associations found in the current study are candidates as LD expands beyond the sequenced region, and could affect gene expression through regulatory elements outside the coding region such as the promoter. The confirmation of associated polymorphisms within candidate genes must be done in separate association mapping populations or through targeted mutation studies. Future studies will include adequate diversity and more lines to increase the statistical power to detect marker trait associations.

References

1. R. Tuberosa, S. Salvi, From QTLs to genes controlling root traits in maize, in: J.H.J. Spiertz, P.C. Struik, H.H. Van Laar HH (Eds.), Scale and complexity in plant systems research: gene–plant–crop relations, Springer, 2007, pp. 13–22.
2. H.A. Mooney, A.J. Bloom, F.S. Chapin III, Resource limitation in plants-an economic analogy. Annu. Rev. Ecol. Syst. 16 (1985) 363-392.
3. D. Robinson, Compensatory changes in the partitioning of dry matter in relation to nitrogen uptake and optimal variations of growth, Ann. Bot.-London 58 (1986) 841-848.
4. D.W. Hilbert, Optimization of plant root: shoot ratios and internal nitrogen concentration. Ann. Bot.-London 66 (1990) 91-99.
5. H.L. Reynolds, C. D'Antonio, The ecological significance of plasticity in root weight ratio in response to nitrogen: Opinion. Plant Soil 185 (1996) 75-97.
6. K.D. Bonifas, D.T. Walters, K.G. Cassman, J.L. Lindquist, Nitrogen supply affects root: shoot ratio in corn and velvetleaf (*Abutilon theophrasti*), Weed Sci. 53 (2005) 670-675.

7. A.H. Abdel-Ghani, B. Kumar, J. Reyes-Matamoros, P.J. Gonzalez-Portilla, C. Jansen, J.P. San Martin, M. Lee, T. Lübberstedt T, Genotypic variation and relationships between seedling and adult plant traits in maize (*Zea mays* L.) inbred lines grown under contrasting nitrogen levels. Euphytica 189 (2013) 123-133.
8. P.H. Nye, P.B. Tinker, Solute movement in the soil-root system, University of California press, Berkeley, 1977.
9. B. Sattelmacher, F. Klotz, H. Marschner, Influence of the nitrogen level on root growth and morphology of two potato varieties differing in nitrogen acquisition, Plant Soil 123 (1990) 131-137.
10. E.A. Hurd, T.F. Townley-Smith, L.A. Patterson, C.H. Owen, Techniques used in producing Wascana wheat, Can. J. Plant Sci. 52 (1972) :689–691.
11. S.A. Barber, A.D. MacKay, Root growth and phosphorus and potassium uptake by two corn genotypes in the field. Fert. Res. 10 (1986) 217–230.
12. P.J. Gregory, Root growth and activity, in: G.A. Peterson (Ed.) Physiology and determination of crop yield, ASA, CSSA, and SSSA, Madison, WI, 1994, pp. 65–93.
13. H. Marschner, Role of root growth, arbuscular mycorrhiza, and root exudates for the efficiency in nutrient acquisition. Field Crop Res. 56 (1988) 203–207.
14. B. Kumar, A.H. Abdel-Ghani, J. Reyes-Matamoros, F. Hochholdinger, T. Lübberstedt, Genotypic variation for root architecture traits in seedlings of maize (*Zea mays* L.) inbred lines. Plant Breeding 131 (2012) 465-478.
15. R.H. Andrew, S.S. Solanki SS, Comparative root morphology for inbred lines of corn as related to performance. Agron. J. 58 (1966) 415-418.

16. M.S. Zuber, Evaluation of corn root system under various environments. 23rd Corn and Sorghum Conf., 1968, pp. 1-19.

17. H.G. Nass, M.S. Zuber, Correlation of corn (*Zea mays* L.) roots to early in development to mature root development, Crop Sci. 11 (1971) 655-658.

18. R. Tuberosa, S. Salvi, M.C. Sanguinetti, M. Maccaferi, S. Giuliani, P. Landi, Searching for quantitative trait loci controlling root traits in maize: a critical appraisal. Plant Soil 255 (2003) 35-54.

19. K. Woll, L. Borsuk, H. Stransky, D. Nettleton, P.S. Schnable, F. Hochholdinger, Isolation, characterization and pericycle specific transcriptome analyses of the novel maize (*Zea mays* L.) lateral and seminal root initiation mutant rum1. Plant Physiol. 139 (2005) 1255-1267.

20. E.C. Abbe, O.L. Stein, The origin of the shoot apex in maize: embryogeny, Am. J. Bot. 41 (1954) 285-293.

21. T.J. Wen, F. Hochholdinger, M. Sauer, W. Bruce, P.S. Schnable, The *roothairless1* gene of maize (*Zea mays*) encodes a homolog of *sec3*, which is involved in polar exocytosis, Plant Physiol. 138 (2005) 1637-1643.

22. **G.** Taramino, M. Sauer, J. Stauffer, D. Multani, X. Niu, H. Sakai, F. Hochholdinger, The *rtcs* gene in maize (*Zea mays* L.) encodes a lob domain protein that is required for postembryonic shoot-borne and embryonic seminal root initiation. Plant J**.** 50 (2007) 649-659.

23. I. Von Behrens, M. Komatsu, Y. Zhang, K.W. Berendzen, X. Niu, H. Sakai, G. Taramino, F. Hochholdinger, Rootless with undetectable meristem1 encodes a

monocot-specific AUX/IAA protein that controls embryonic seminal and postembryonic lateral root initiation in maize. Plant J. 66 (2011) 341-353.

24. W. Hertz, F. Hochholdinger, F. Schwall, G. Feix, Isolation and characterization of rtcs, a mutant deficient in the formation of nodal roots, Plant J. 10 (1996) 845-857.
25. F. Hochholdinger, T.J. Wen, R. Zimmermann, P. Chimot-Marolle, O. da Costa e Silva, W. Bruce, K.R. Lamkey, U. Wienand, P.S. Schnable, The maize (*Zea mays* L.) roothairless3 gene encodes a putative GPI-anchored, monocot-specific, COBRA-like protein that significantly affects grain yield, Plant J. 54 (2008) 888.
26. S.M. Brady, S. Song, K.S. Dhugga, J.A. Rafalski, P.N. Benfey, Combining Expression and Comparative Evolutionary Analysis. The COBRA Gene Family. *Plant Physiol.* 143 (2007) 172-187.
27. K.B. Alpert, S.D. Tanksley, High-resolution mapping and isolation of a yeast artificial chromosome contig containing fw2.2: a major fruit weight quantitative trait locus in tomato, Proc. Natl. Acad. Sci. USA 93 (1996) 15503–15507.
28. C.W. Stuber, M. Polacco, M.L. Senior, Synergy of empirical breeding, marker-assisted selection, and genomics to increase crop yield potential. Crop Sci. 39 (1999) 1571-1583.
29. J. Jannink, A.J. Lorenz, H. Iwata, Genomic selection in plant breeding: from theory to practice, Brief Funct. Genomics 9 (2010) 166-177.
30. J.M. Thornsberry, M.M. Goodman, J. Doebley, S. Kresovich, D. Nielsen, E.S. Buckler, Dwarf8 polymorphisms associate with variation in flowering time, Nat. Genet. 28 (2001) 286-289.

31. J. Yu, E.S. Buckler, Genetic association mapping and genome organization of maize. Curr. Opin. Biotech. 17 (2006) 155–160.

32. P.K. Gupta, S. Rustgi, P.L. Kulwal, Linkage disequilibrium and association studies in higher plants: Present status and future prospects. Plant Mol. Biol. 57 (2005) 461-485.

33. J.R. Andersen, T. Lübberstedt, Functional markers in plants, Trends Plant Sci. 8 (2003) 554 –560.

34. K.A. Palaisa, M. Morgante, M. Williams, A. Rafalski, Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci, Plant Cell 15 (2003) 1795-1806.

35. J.R. Andersen, I. Zein, G. Wenzel, B. Krützfeldt, J. Eder, M. Ouzunova, T. Lübberstedt, High levels of linkage disequilibrium and associations with forage quality at a phenylalanine ammonia-lyase locus in European maize (*Zea mays* L.) inbreds, Theor. Appl. Genet. 114 (2007) 307-19.

36. A. Brenner, I. Zein, Y. Chen, J.R. Andersen, G. Wenzel, M. Ouzunova, J. Eder, B. Darnhofer, U. Frei, Y. Barrière, T. Lübberstedt, Polymorphisms in O-methyltransferase genes are associated with stover cell wall digestibility in European maize (*Zea mays* L.), BMC Plant Biol. 10 (2010) 27.

37. Y. Chen, I. Zein, E.A. Brenner, J.R. Andersen, M. Landbeck, M. Ouzunova, T. Lübberstedt, Polymorphisms in monolignol biosynthetic genes are associated with biomass yield and agronomic traits in European maize (*Zea mays* L.), BMC Plant Biol. 10 (2010) 12.

38. T.L. Setter, J. Yan, M. Warburton, J.M. Ribaut, Y. Xu, M. Sawkins, E.S. Buckler, Z. Zhang, M.A. Gore, Genetic association mapping identifies single nucleotide

polymorphisms in genes that affect abscisic acid levels in maize floral tissues during drought, J. Exp. Bot. 62 (2011) 701–716.

39. A.M. Hallauer, J.B. Miranda, Quantitative genetics in maize breeding 2nd edition, Iowa State University Press, Ames, Iowa, 1981.
40. J. Rozas, J.C. Sanchez-DelBarrio, X. Messeguer, R. Rozas, DnaSP DNA polymorphism analyses by the coalescent and other methods, Bioinformatics 19 (2003) 2496-2497.
41. M. Nei, Molecular evolutionary genetics, Columbia University Press, New York, NY, 1987.
42. G.A. Watterson, On the number of segregating sites in genetical models without recombination, Theor. Popul. Biol. 7 (1975) 256-276.
43. Y. Li, G. Haseneyer, C. Schön, D. Ankerst, V. Korzun, P. Wilde, E. Bauer, High levels of nucleotide diversity and fast decline of linkage disequilibrium in rye (*Secale cereale* L.) genes involved in frost response, BMC Plant Biol. 11 (2011) 6.
44. F. Tajima, Evolutionary relationship of DNA sequences in finite populations, Genet 105 (1983) 437-460.
45. F. Tajima, DNA Polymorphism in a subdivided population: the expected number of segregating sites in the two-subpopulation model, Genetics 123 (1989) 229-240.
46. W.B. Barbazuk, S.E. Emrich, H.D. Chen, L. Li, P.S. Schnable, SNP discovery via 454 transcriptome sequencing. Plant J. 51 (2007) 910-918.
47. J. K. Pritchard, M. Stephens, P. Donnelly, Inference of Population Structure Using Multilocus Genotype Data, Genetics155 (2000) 945-959.

48. G. Evano, S. Regnaut, J. Goudet, Detecting the number of clusters of individuals using the software structure: a simulation study, Mol. Ecol. 14 (2005) 2611-2620.

49. P.J. Bradbury, Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, E.S. Buckler, TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23 (2007) 2633-2635.

50. Z. Zhang, E. Ersoz, C. Lai, R.J. Todhunter, H.K. Tiwari, M.A. Gore, P.J. Bradbury, J. Yu, D.K. Arnett, J.M. Ordovas, E.S. Buckler,Mixed linear model approach adapted for genome wide association studies. Nat. Genet. 42 (2010) 355-60.

51. C.R. Henderson, Best linear unbiased estimation and prediction under a selection model, Biometrics 31 (1975) 423–447.

52. J.D. Storey, J.E. Taylor, D. Siegmund, Strong control, conservative point estimation, and simultaneous conservative consistency of false discovery rates: A unified approach, J. Roy. Stat. Soc. B. 66 (2004) 187-205.

53. K. Higo, Y. Ugawa, M. Iwamoto, T. Korenaga, Plant cis-acting regulatory DNA elements (PLACE) database, Nucleic Acids Res. 27 (1999) 297-300.

54. B. Kumar, Candidate gene based association study for nitrogen use efficiency and associated traits in maize; Ph.D. Thesis (2013) Iowa State University, Ames, Iowa

55. J.R. Andersen, I. Zein, G. Wenzel, B. Darnhofer, J. Eder, M. Ouzunova, T. Lübberstedt, Characterization of phenylpropanoid pathway genes within European maize (*Zea mays* L.) inbreds, BMC Plant Biol. 8 (2008) 2.

56. I. Zein, G. Wenzel, J.R. Andersen, T. Lübberstedt, Low Level of Linkage Disequilibrium at the COMT (Caffeic Acid O-methyl Transferase) Locus in European Maize (*Zea mays* L.). Genet. Resour. Crop Ev. 54 (2007) 139-148.

57. Y. Xing, U Frei, B. Schejbel, T. Asp, T. Lübberstedt, Nucleotide diversity and linkage disequilibrium in 11 expressed resistance candidate genes in *Lolium perenne*, BMC Plant Biol. 7 (2007) 43.

58. E. Guingo, Y. Hébert, A. Charcosset, Genetic analysis of root traits in maize, Agronomie 18 (1988) 225–235.

59. A.Y. Kamara, J.G. Kling, S.O. Ajala, A. Menkir, The relationship between vertical root-pulling resistance and nitrogen uptake and utilization in maize breeding lines, Maydica 47 (2002) 135-140.

60. M. Coque, A. Gallais, Genomic regions involved in response to grain yield selection at high and low nitrogen fertilization in maize, Theor. Appl. Genet. 112 (2006) 1205–1220.

61. C. Majer, F. Hochholdinger, Defining the boundaries: structure and function of LOB-domain proteins, Trends Plant Sci. 16 (2011) 47-52.

62. H. Iwakawa, Y. Ueno, E. Semiarti, H. Onouchi, S. Kojima, H. Tsukaya, M. Hasebe, T. Soma, M. Ikezaki, C. Machida, Y. Machida,: The ASYMMETRIC LEAVES2 gene of Arabidopsis thaliana, required for formation of a symmetric flat leaf lamina, encodes a member of a novel family of proteins characterized by cysteine repeats and a leucine zipper, Plant Cell Physiol. 43 (2002) 467–478.

63. H. Liu, S. Wang, X. Yu, J. Yu, X. He, S. Zhang, H. Shou, P. Wu,ARL1, a LOB-domain protein required for adventitious root formation in rice, Plant J. 43 (2005) 47-56.

64. Y. Inukai, T. Sakamoto, M. Ueguchi-Tanaka, Y. Shibata, K. Gomi, L. Umemura, Y. Hasegawa, M. Ashikari, H. Kitano, M. Matsuokaa, Crown rootless1, Which Is

Essential for Crown Root Formation in Rice, Is a Target of an AUXIN RESPONSE FACTOR in Auxin Signaling, Plant Cell 17 (2005) 1387-1396.

65. T. Vamerali, M. Saccomani, S. Bona, G. Mosca, M. Guarise, A. Ganis, A comparison of root characteristics in relation to nutrient and water stress in two maize hybrids. Plant soil 255 (2003) 157-167.

Table 1 *Rtcl* haplotypes formed by 45 single nucleotide polymorphisms and average phenotypic values of lines included in the individual haplotypes.

SNP position	Exon 1																Intron 1							Exon 2																					
	33	94	204	232	290	296	298	307	317	320	324	332	357	373	374	413	468	473	481	510	531	543	547	554	576	597	601	607	616	632	648	694	695	703	704	708	711	716	720	736	737	759	761	799	824
	s	a	s	a	a	a	a	a	a	a	s	a	s	a	a	a								s	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	s	a	a
Haplotypes																																													
Hap_1	G	C	C	C	T	A	A	T	C	C	G	C	A	G	G	C	A	A	G	A	G	A	T	C	C	G	G	N	C	G	A	G	G	C	G	A	A	C	T	A	T	C	T	T	A
Hap_2	G	C	C	C	T	A	A	T	C	C	C	C	A	G	A	C	C	G	A	T	G	A	T	C	C	G	G	N	C	G	A	G	G	C	G	A	A	C	A	C	T	C	T	G	A
Hap_3	G	C	C	C	T	A	A	T	T	C	C	A	A	G	A	C	A	A	G	A	G	A	T	C	C	G	A	N	T	G	T	G	T	C	G	T	A	A	A	N	A	C	C	G	C
Hap_4	A	G	A	C	T	A	A	T	C	T	G	A	A	G	G	C	A	A	G	A	G	A	T	C	C	G	G	N	C	G	A	G	G	C	G	A	A	C	T	A	T	C	T	T	A
Hap_5	A	G	C	C	A	G	T	T	T	C	T	C	T	A	N	A	T	G	G	T	C	G	G	T	C	T	G	N	C	A	A	G	G	G	G	A	A	C	A	C	T	C	T	A	C
Hap_6	G	C	C	C	A	G	T	T	T	C	T	C	T	A	N	A	T	G	G	A	G	A	T	C	C	G	G	N	C	G	A	G	G	C	G	A	A	C	A	C	T	C	T	G	A
Hap_7	G	C	C	C	T	A	A	T	C	C	C	C	A	G	A	C	A	A	A	T	G	A	T	T	C	G	G	N	C	G	G	G	G	C	G	A	A	C	A	N	A	G	T	T	C
Hap_8	G	C	C	A	T	A	A	T	T	C	C	C	T	A	N	C	A	A	A	T	G	A	T	T	C	G	G	N	C	G	G	G	G	C	G	A	A	C	A	N	A	G	T	T	C
Hap_9	G	C	C	C	T	A	A	T	C	C	C	C	A	G	A	C	N	A	A	T	G	A	T	T	C	G	G	N	C	G	G	G	G	C	G	A	A	C	A	N	A	G	T	T	C
Hap_10	G	C	A	C	T	A	A	T	C	T	G	A	A	G	G	C	A	A	G	T	G	A	T	C	C	G	G	N	C	G	G	A	G	N	A	T	T	C	A	N	A	C	T	T	C
Hap_11	G	C	C	C	T	A	A	T	C	T	G	A	G	G	G	C	A	A	G	T	G	A	T	C	C	G	G	N	C	G	G	A	G	N	A	T	T	C	A	N	A	C	T	T	C
Hap_12	G	C	A	A	T	A	A	A	C	C	C	C	A	G	G	A	T	G	G	T	G	A	G	T	T	G	G	G	C	G	A	A	G	N	A	T	T	C	A	N	A	C	T	T	C
Hap_13	G	C	C	C	T	A	A	T	T	C	C	A	A	G	A	C	A	A	G	A	G	A	T	C	C	G	A	C	T	G	T	G	T	C	G	T	A	A	A	N	A	C	C	G	C
Hap_14	G	C	C	C	T	A	A	T	T	C	C	A	A	G	A	C	A	A	G	A	G	A	T	C	C	G	A	N	T	G	T	G	T	C	G	T	A	N	A	N	A	C	C	G	C
Hap_15	A	G	C	C	A	G	T	T	T	C	T	C	T	A	N	A	T	G	G	T	G	G	G	T	C	T	G	N	C	A	A	G	G	G	G	A	A	C	A	C	T	C	T	A	C
Hap_16	G	G	C	C	T	A	A	T	T	C	C	A	A	G	A	C	A	A	G	T	G	G	G	T	C	T	G	N	C	A	A	G	G	G	G	A	A	C	A	C	T	C	T	A	C

Haplotypes	sTRL	tTRL	fTRL	sRDW	tRDW	fRDW
Hap_1	32.4	135.9	219.3	12.40	29.93	47.88
Hap_2	43.4	121.6	207.9	14.61	30.00	47.79
Hap_3	36.7	131.2	216.8	15.92	33.16	51.59
Hap_4	33.7	134.7	247.6	13.23	27.54	51.99
Hap_5	47.4	155.7	246.0	16.43	33.97	52.80
Hap_6	41.2	160.0	195.3	12.78	32.89	37.70
Hap_7	34.8	145.5	236.1	12.09	29.88	47.17
Hap_8	45.0	127.9	256.2	13.31	32.35	54.25
Hap_9	29.2	109.8	231.5	9.77	24.37	52.20
Hap_10	21.8	117.5	201.3	6.46	16.11	26.77
Hap_11	27.2	156.9	268.4	9.37	26.84	45.50
Hap_12	48.4	153.1	250.9	14.45	31.81	52.49
Hap_13	38.8	158.4	251.9	16.17	35.07	56.27
Hap_14	43.0	153.6	235.4	17.12	38.07	60.56
Hap_15	50.6	119.9	210.6	19.31	34.45	60.00
Hap_16	43.2	147.4	205.7	14.56	32.47	45.32
Maximum	50.6	160.0	268.4	19.31	38.07	60.56
Minimum	21.8	109.8	195.3	6.46	16.11	26.77
Range	28.8	50.2	73.1	12.85	21.96	33.79

s = synonymous substitution: a = non-synonymous substitution resulting in amino acid change; N = missing nucleotide; sTRL = Total root length at 6th day; tTRL = Total root length at 10th day; fTRL = Total root length at 14th day; sRDW = Root dry weight at 6th day; tRDW = Root dry weight at 10th day; fRDW = Root dry weight at 14th day.

Table 2
Rth3 haplotypes formed by 15 single nucleotide polymorphisms and average phenotypic values of lines included in the individual haplotypes.

SNP Position																				
								Exon												
	5	1	4	1	2	3	3	3	4	4	4	4	5	6	6					
	9	6	7	8	3	5	9	9	1	3	6	9	1	0	2					
		3	9	0	4	1	3	9	7	8	5	2	9	0	1					
	a	a	a	s	s	s	s	s	s	s	s	s	s	s	s					
Haplotypes																sTRL	tTRL	fTRL	sRDW	tRDW	fRDW
Hap_1	C	G	G	C	G	G	G	C	G	C	A	G	G	T	G	39.9	128.8	215.6	14.74	34.13	50.32
Hap_2	C	A	G	C	G	G	G	G	A	C	A	G	G	T	G	42.7	144.9	235.6	15.96	32.69	52.23
Hap_3	N	N	G	C	G	G	G	G	A	C	A	G	G	T	G	40.9	134.8	278.9	16.51	36.29	62.89
Hap_4	C	G	T	T	G	G	G	C	G	A	A	T	A	C	A	34.1	130.2	204.0	11.26	28.21	42.32
Hap_5	C	G	G	C	T	A	A	C	A	C	A	G	G	T	A	32.1	143.8	250.4	11.88	29.38	52.05
Hap_6	C	G	G	C	G	G	G	G	A	C	A	G	G	T	G	35.9	141.3	224.9	12.89	29.67	48.22
Hap_7	T	G	G	C	G	G	G	G	A	C	A	G	G	T	G	45.7	150.5	230.0	16.69	33.23	50.20
Hap_8	N	G	G	C	G	G	G	G	A	C	A	G	G	T	G	20.1	149.1	194.2	12.41	29.99	40.60
Hap_9	C	A	G	C	G	G	G	G	A	C	G	G	G	T	G	46.3	143.7	247.0	14.64	33.01	58.76
													Maximum			46.3	150.5	278.9	16.69	36.29	62.89
													Minimum			20.1	128.8	194.2	11.26	28.21	40.60
													Range			26.2	21.7	84.7	5.43	8.08	22.29

s = synonymous substitution: a = non-synonymous substitution resulting in amino acid change; N = missing nucleotide; sTRL = Total root length at 6th day; tTRL = Total root length at 10th day; fTRL = Total root length at 14th day; sRDW = Root dry weight at 6th day; tRDW = Root dry weight at 10th day; fRDW = Root dry weight at 14th day.

Table 3
Rum1 haplotypes formed by 12 single nucleotide polymorphisms and average phenotypic values of lines included in the individual haplotypes.

SNP position																		
	Exon 5							Intron 4										
	63	78	118	236	251	264	276	302	358	381	405	407						
	s	s	a															
Haplotypes													sTRL	tTRL	fTRL	sRDW	tRDW	fRDW
Hap_1	T	C	G	A	A	C	T	T	G	T	C	A	35.3	121.3	199.1	11.47	25.44	40.89
Hap_2	C	C	G	A	T	A	A	G	G	T	T	G	44.8	150.0	240.8	15.81	33.85	53.11
Hap_3	C	C	C	A	T	C	T	T	C	C	C	A	42.4	155.7	251.2	15.91	36.02	58.57
Hap_4	N	C	G	N	T	A	T	N	G	N	N	N	31.1	119.6	170.0	16.98	33.15	48.05
Hap_5	N	C	N	A	T	A	A	G	G	N	N	N	68.0	159.7	242.5	28.11	41.79	56.38
Hap_6	N	C	C	A	T	A	T	T	G	N	N	N	32.1	140.6	297.3	18.22	36.90	69.42
Hap_7	N	C	C	A	T	N	N	N	G	N	N	N	34.2	140.5	226.2	14.11	31.62	47.4
Hap_8	C	C	C	A	T	N	N	T	C	C	C	A	28.1	158.6	251.3	15.89	34.71	52.7
Hap_9	C	C	C	A	T	C	T	T	C	N	N	N	29.0	102.2	194.5	16.08	35.65	49.19
Hap_10	N	C	G	A	T	A	A	A	G	N	N	N	20.8	80.4	113.6	6.71	17.54	21.92
Hap_11	N	T	G	N	T	N	T	N	G	N	N	N	29.2	152.7	245.7	10.65	29.20	43.48
Hap_12	C	C	C	A	T	N	T	T	C	N	N	N	38.2	123.9	200.6	21.24	39.82	61.69
Hap_13	T	C	G	A	A	C	T	T	G	N	N	N	35.1	131.1	198.2	13.29	29.50	43.87
Hap_14	N	C	G	A	T	A	A	N	G	N	N	N	36.4	147.0	252.1	10.45	26.40	47.35
Hap_15	C	C	G	A	A	C	T	T	G	T	C	A	41.2	160.0	195.3	12.78	32.89	37.7
Hap_16	C	C	G	G	T	C	T	T	G	T	C	A	41.1	145.5	249.2	15.21	35.20	57.61
Hap_17	T	T	G	G	T	C	T	T	G	T	T	G	38.0	147.1	235.0	13.40	30.68	50.37
Hap_18	C	C	G	A	T	A	A	T	G	T	T	G	20.3	107.1	238.4	7.71	23.34	38.11
Hap_19	T	T	G	A	A	C	T	T	G	T	C	A	24.2	129.0	178.8	11.29	34.85	42.44
Hap_20	N	C	G	A	A	C	T	T	G	N	N	N	45.1	132.7	246.1	15.01	26.55	59.29
Hap_21	N	C	C	A	T	N	T	N	C	N	N	N	30.5	122.1	266.9	10.77	21.87	50.39
Hap_22	C	C	G	G	T	C	T	T	G	T	T	G	52.0	156.8	221.1	16.95	31.19	49.09
													Maximum	68	160	297.3	28.11	41.79	69.42
													Minimum	20.3	80.4	113.6	6.71	17.54	21.92
													Range	47.7	79.6	183.7	21.4	24.25	47.5

s = synonymous substitution: a = non-synonymous substitution resulting in amino acid change; N = missing nucleotide; sTRL = Total root length at 6th day; tTRL = Total root length at 10th day; fTRL = Total root length at 14th day; sRDW = Root dry weight at 6th day; tRDW = Root dry weight at 10th day; fRDW = Root dry weight at 14th day.

Table 4
Rul1 haplotypes formed by 6 single nucleotide polymorphisms and average phenotypic values of lines included in the individual haplotypes.

SNP position												
			Exon 6			Intron5						
	2	7	1	2	2	4						
	2	5	0	2	4	0						
			1	6	4	4						
	a	s	a	a	a							
Haplotypes							sTRL	tTRL	fTRL	sRDW	tRDW	fRDW
Hap_1	A	C	G	C	C	T	39.7	140.1	228.5	13.71	30.67	49.62
Hap_2	A	C	G	C	C	G	37.4	114.2	159.7	12.84	24.76	34.48
Hap_3	A	C	G	A	T	A	40.1	146.7	239.5	15.49	34.03	54.09
Hap_4	G	T	A	C	C	G	38.7	132.1	207.1	15.40	28.81	42.36
Hap_5	A	C	G	C	T	G	41.2	152.8	222.3	14.00	34.05	47.66
Hap_6	N	C	G	A	T	A	27.2	156.9	268.4	9.37	26.84	45.5
Hap_7	G	T	A	C	C	T	30.5	122.1	266.9	10.77	21.87	50.39
				Maximum			41.2	156.9	268.4	15.49	34.05	54.09
				Minimum			27.2	114.2	159.7	9.37	21.87	34.48
				Range			14.0	42.7	108.7	6.12	12.18	19.61

s = synonymous substitution: a = non-synonymous substitution resulting in amino acid change; N = missing nucleotide; sTRL = Total root length at 6th day; tTRL = Total root length at 10th day; fTRL = Total root length at 14th day; sRDW = Root dry weight at 6th day; tRDW = Root dry weight at 10th day; fRDW = Root dry weight at 14th day.

Table 5
Summary of alignment length, number of genotypes per alignment, polymorphisms and nucleotide diversity in the *Rtcl*, *Rth3*, *Rum1* and *Rul1* genes in maize.

	Entire region	Non-coding region	Coding region			No. of Haplotypes	*Hd*
			All sites	Synonymous	Non-synonymous		
Rtcl (n=69)	830bp					16	0.873
No. of segregating sites	45	7	38	6	32		
π	0.021	0.022	0.021	0.017	0.017		
Tajima's D	2.691**	2.232*	2.569*	2.593*	2.278*		
Rth3 (n=74)	713bp					9	0.786
No. of segregating sites	15	0	15	12	3		
π	0.007	n.a	0.007	0.026	0.0002		
Tajima's D	1.298NS	n.a	1.298NS	1.500NS	(-)0.605NS		
Rum1 (n=74)	432bp					22	0.855
No. of segregating sites	12	9	3	2	1		
π	0.011	0.017	0.005	0	0		
Tajima's D	0.960^{NS}	1.033^{NS}	0.306^{NS}	n.a	n.a		
Rul1 (n=74)	411bp					7	0.624
No. of segregating sites	6	1	5	1	4		
π	0.005	0.007	0.004	0.003	0.005		
Tajima's D	1.766^{NS}	2.305*	1.073^{NS}	$(-)0.322^{NS}$	1.465^{NS}		

Numbers of lines are shown in the parenthesis. ns = not significant;*p<0.05; **p<0.01.

Table 6
Polymorphic sites of *Rtcl* gene associated with the root traits (Total root length, Root dry weight) at different growth stages identified by MLM analysis.

Site	SNP	Amino acid change	E/I	Days of measurement		
				Six	Ten	Fourteen
290	T→A	Leu-His	E1	TRL:RDW	-	-
296	A→G	Asp-Gly	E1	TRL:RDW	-	-
298	A→T	Ser-Cys	E1	TRL:RDW	-	-
317	C→T	Pro-Leu	E1	TRL:RDW	-	-
320	T→C	Val-Ala	E1	RDW	RDW	-
324	G/C/T	Syn	E1	TRL:RDW	-	-
357	G/A/T	Syn	E1	TRL	-	-
373	G→A	Asp-Asn	E1	TRL	-	-
374	A→G	Asp-Gly	E1	RDW	RDW	-
413	C→A	Thr-Lys	E1	TRL	TRL	-
468	A→T	-	I1	TRL:RDW	-	-
473	A→G	-	I1	TRL:RDW	TRL	-
510	T→A	-	I1	TRL	-	TRL
531	G→C	-	I1	TRL	TRL	-
543	G→A	-	I1	TRL:RDW	-	-
547	T→G	-	I1	TRL:RDW	TRL	-
554	C→T	Ala-Val	E2	TRL	TRL	TRL
597	G→T	Syn	E2	TRL:RDW	-	-
632	A→G	Glu-Gly	E2	TRL:RDW	-	-
703	C→G	Arg-Gly	E2	TRL:RDW	-	-
720	A→T	Syn	E2	TRL	-	-
736	C→A	His-Asn	E2	TRL:RDW	-	-
799	T→G	Trp-Gly	E2	TRL:RDW	-	-

TRL=Total Root Length; RDW = Root Dry Weight

Table 7
Polymorphic sites of *Rth3* gene associated with the root traits (Total root length, Root dry weight) at different growth stages identified by MLM analysis**.**

Site	SNP	Amino acid change	E/I	Days of measurement		
				Six	Ten	Fourteen
163	G→A	Ala-Thr	E	TRL	-	RDW
180	G→T	Syn	E	TRL	-	TRL; RDW
234	C→T	Syn	E	TRL	-	TRL; RDW
351	G→T	Syn	E	TRL	-	
393	G→A	Syn	E	TRL	-	
399	G→A	Syn	E	TRL	-	
417	G→C	Syn	E	TRL	-	
438	A→G	Syn	E	TRL	-	TRL; RDW
465	C→A	Syn	E	TRL	-	TRL; RDW
492	G→T	Syn	E	TRL	-	TRL; RDW
519	G→A	Syn	E	TRL	-	TRL; RDW
600	T→C	Syn	E	TRL	-	TRL; RDW
621	G→A	Syn	E	TRL;RDW	-	

TRL=Total Root Length; RDW = Root Dry Weight

Table 8
Polymorphic sites of *Rum1* gene associated with the root traits (Total root length, Root dry weight) at different growth stages identified by MLM analysis.

Site	SNP	Amino acid change	E/I	Days of measurement		
				Six	Ten	Fourteen
63	C→T	Val-Ala	E5	-	TRL;RDW	TRL;RDW
118	G→C	Val-Leu	E5	RDW	-	RDW
251	T→A	-	I4	-	TRL;RDW	TRL;RDW
302	T→G	-	I4	TRL;RDW	-	TRL;RDW
358	G→C	-	I4	-	-	RDW
381	T→C	-	I4	-	-	RDW

TRL=Total Root Length; RDW = Root Dry Weight

Table 9
Polymorphic sites of *Rul1* gene associated with the root traits (Total root length, Root dry weight) at different growth stages identified by MLM analysis.

Site	SNP	Amino acid change	E/I	Days of measurement		
				Six	Ten	Fourteen
7	T→A→G	-	I5	-	-	RDW
311	G→A	Syn	E6	-	RDW	-
336	C→T	Thr-Ile	E6	-	RDW	-
389	A→G	Ser-Gly	E6	-	RDW	-

RDW = Root Dry Weight

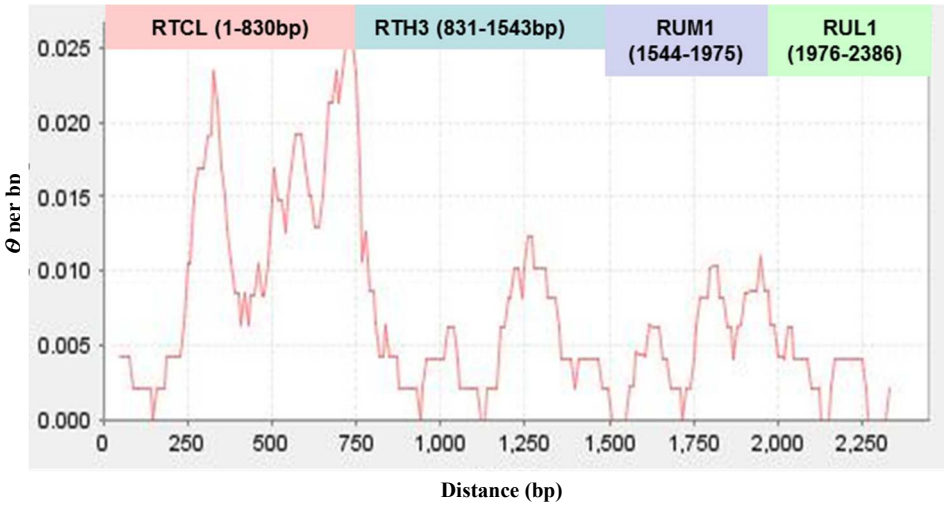
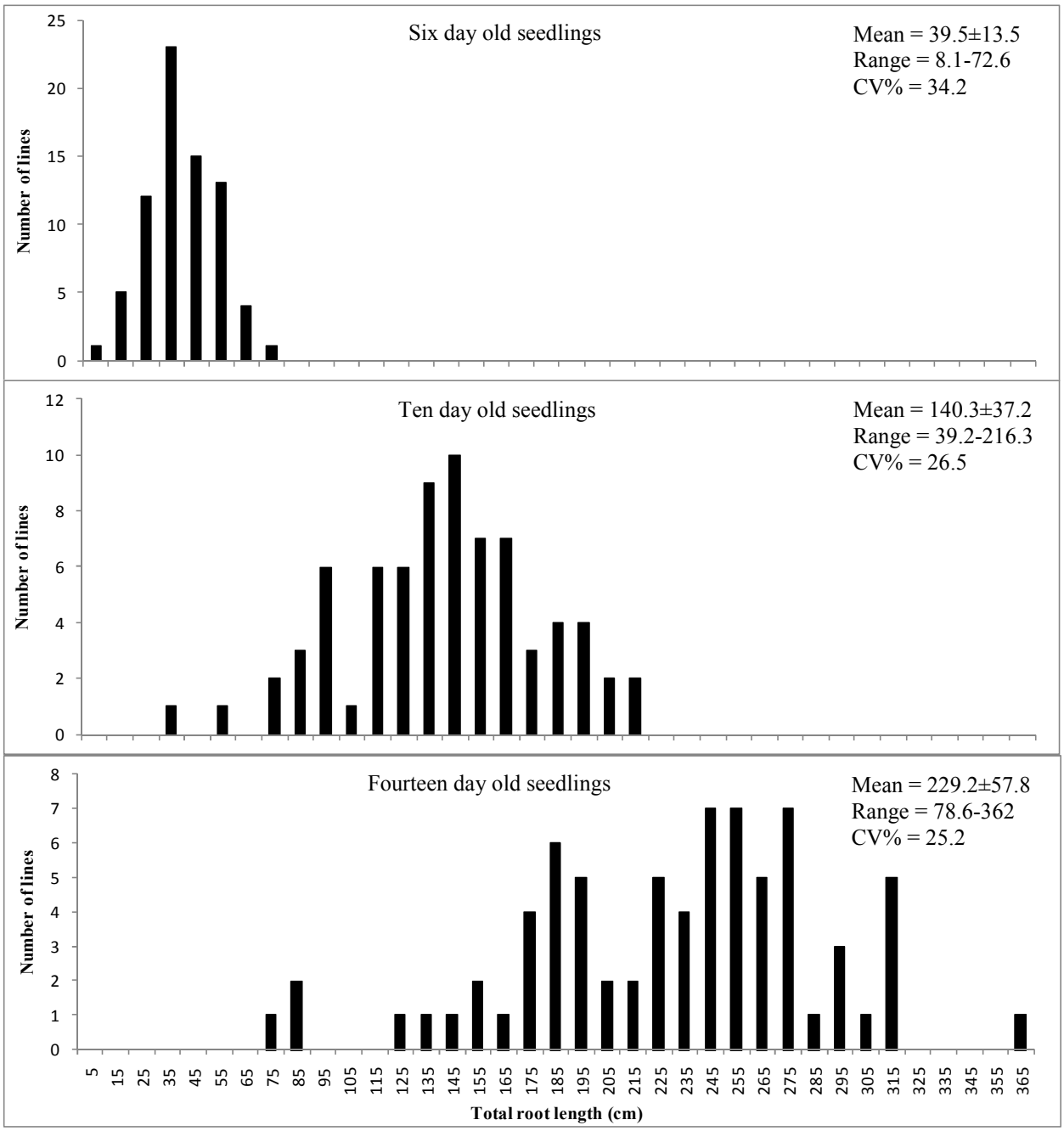
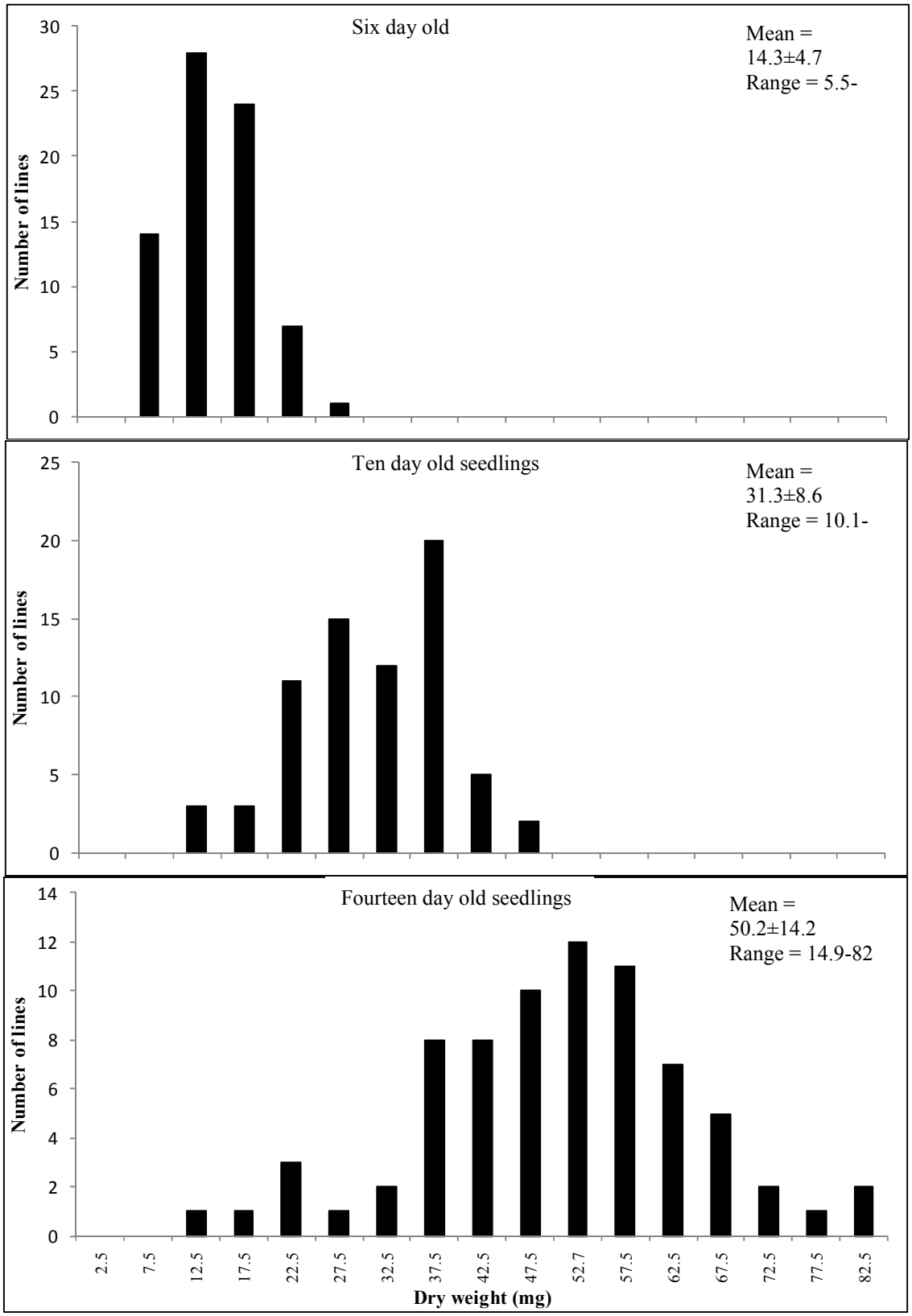


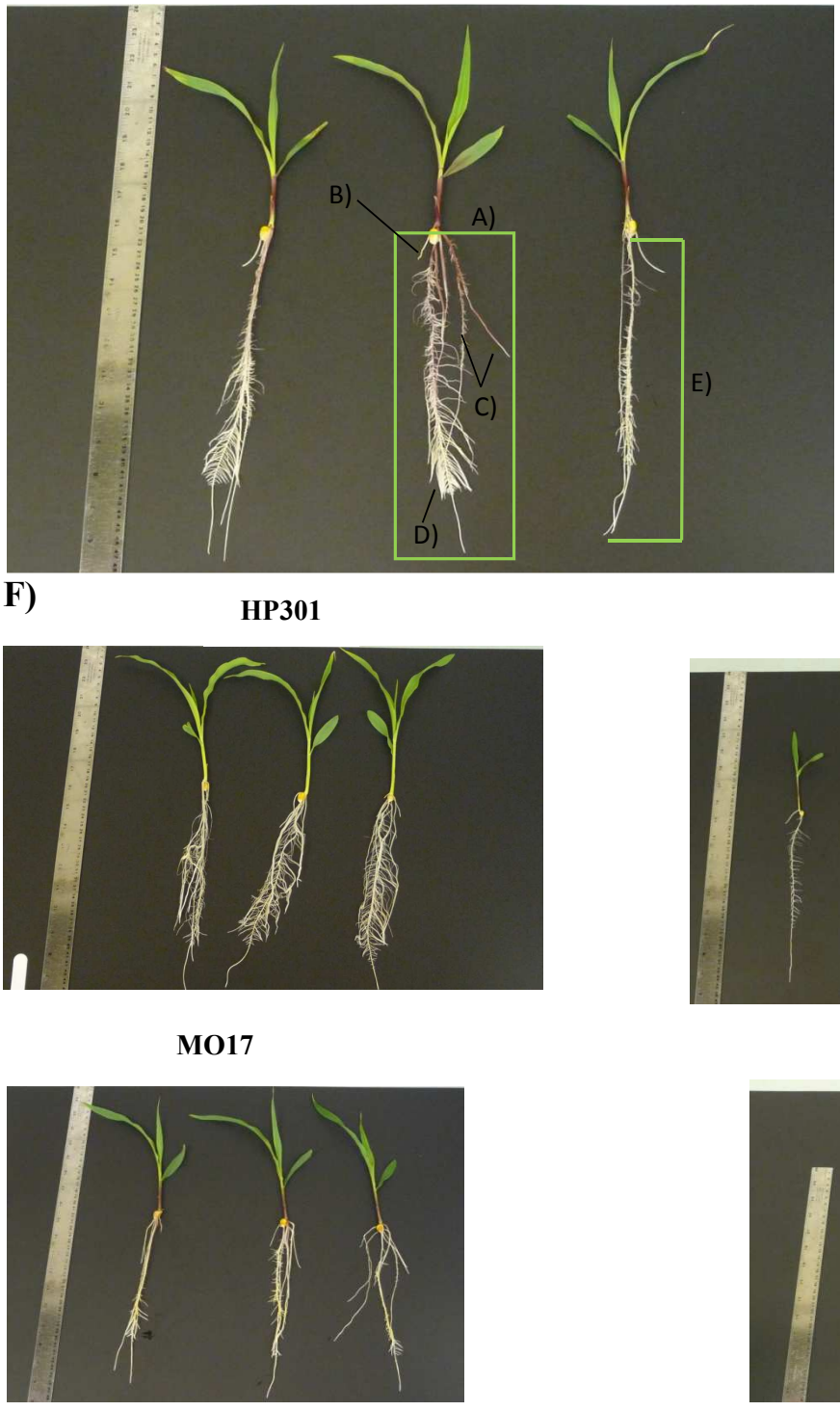
Fig. 1. Nucleotide diversity values (x-axis) in sliding windows (size = 10, length = 100) at the *Rtcl, Rth3, Rum1* and *Rul1* gene locus for all lines. Nucleotide diversity was calculated based on the θ, neutral mutation parameter derived from the total number of segregating sites. 1-420bp: Rtcl Exon 1; 421-549bp: Rtcl Intron; 550-830bp: Rtcl Exon 2; 831-1543bp: Rth3 Exon; 1544-1750bp: Rum1 Exon 5; 1751-1975bp: Rum1 Intron 4; 1976-2302bp: Rul1 Exon 6; 2303-2386bp: Rul1 Intron 5.



Suppl. Fig 1. Distribution of trait total root length in maize inbred lines, values in x-axis represents mid-point of class-interval.



Suppl. Fig 2. Distribution of trait dry weight in maize inbred lines, values in x-axis represents mid-point of class-interval.



Supl Figure 3. A) Total Root Length, estimated by summing the lengths of primary root, crown, seminal, and lateral roots for each seedling. **B)** Crown Root Number, estimated by counting the number of crown roots on each seedling. **C)** Seminal roots, measured by counting each root and calculated length by measuring each with a ruler and adding lengths for total seminal root length. **D)** Lateral Roots, measured by analyzing the canned images using WinRhizo software. **E)** Primary Root Length, measured manually using a ruler. **F)** Root photos show trait variation regarding length and amount of lateral roots, which make up a major portion of the total root length of the seedlings. These images exemplify the vast amount of variation between the lines studied within this association mapping population.

CHAPTER THREE

ANALYSIS OF MAIZE (*ZEA* MAYS L.) SEEDLING ROOTS WITH THE HIGH-THROUGHPUT IMAGE ANALYSIS TOOL *ARIA* (AUTOMATIC ROOT IMAGE ANALYSIS).

Jordon Pace, Nigel Lee, Hsiang Sing Naik, Baskar Ganapathysubramanian, Thomas Lübberstedt

Paper published in PLoS ONE journal. Abstract, structure, and references are formatted according to the journal standards.

Abstract:

The maize root system is crucial for plant establishment as well as water and nutrient uptake. There is substantial genetic and phenotypic variation for root architecture, which gives opportunity for selection. Root traits, however, have not been used as selection criterion mainly due to the difficulty in measuring them, as well as their quantitative mode of inheritance. Seedling root traits offer an opportunity to study multiple individuals and to enable repeated measurements per year as compared to adult root phenotyping. We developed a new software framework to capture various traits from a single image of seedling roots. This framework is based on the mathematical notion of converting images of roots into an equivalent graph. This allows automated querying of multiple traits simply as graph operations. This framework is furthermore extendable to 3D tomography image data. In order to evaluate this tool, a subset of the 384 inbred lines from the Ames panel, for which extensive genotype by sequencing data are available, was investigated. A genome wide association study was applied to this panel for two traits, Total Root Length and Total Surface Area, captured from seedling root images from WinRhizo Pro 9.0 and the current framework (called *ARIA*) for comparison using 135,311 single nucleotide polymorphism markers. The trait Total Root Length was found to have significant

SNPs in similar regions of the genome when analyzed by both programs. This high-throughput trait capture software system allows for large phenotyping experiments and can help to establish relationships between developmental stages between seedling and adult traits in the future.

Introduction:

The maize (Zea mays L.) root is designed to provide anchorage as well as to secure uptake of water and nutrients, including nitrogen (N), in an efficient manner [1,2]. Maize roots are formed partly during embryonic and partly during post-embryonic development [3]. There are five main types of roots in maize: crown, seminal, primary, lateral, and brace roots [4]. The major portion of root biomass of mature plants is derived from postembryonic, shoot-borne roots. These postembryonic roots include crown roots, formed below soil surface, and brace roots, formed above soil surface [5]. Their function is important to plant performance as they are responsible for the majority of water and nutrient uptake in maize [5].

Two to three week old seedling root systems are made up of primary roots, lateral roots, seminal roots, and root hairs [4,6]. Lateral roots branch outward from the primary root. These root types are called the axial roots and determine root architecture. Lateral roots increase the surface area of the root system and all root types contribute to water and nutrient uptake [2,7,8]. Moreover, lateral roots contain root initiation points, leading to secondary, tertiary, and higher order root structures, with major influence on the overall root architecture of the root stock [2].

There is extensive genetic variation in root architecture. However, root traits have not been considered by plant breeders to select for improved nutrient uptake efficiency or yield improvement due to the difficulty in measuring root traits and their quantitative mode of inheritance [9]. Studying adult roots using maize “shovelomics”, a high-throughput phenotyping

technique that measures adult root traits, is time consuming and laborious. This method of phenotyping is also destructive because roots are dug out of the ground. This limits the number of experiments that can be completed in a season [10]. Changes in maize root architecture may strongly affect yield [11]. Seminal roots play a key role in the acquisition of immobile and mobile nutrients such as phosphorus and nitrogen respectively and can determine spatial and temporal domains of its environment and inter-root competition [6]. The relationship between seminal root biomass in hydroponics and root lodging in a field study focusing on root strength and pulling resistance has been explored. Respective correlations were low, but statistically significant. Correlations found in hydroponic seedling root traits compared to adult field traits were r=0.44* for shoot weight and adult plant height, and r=0.22* for lateral root length with brace root development [12,13].

Seedling phenotyping takes less time, is less laborious, and can be repeated many times during the year allowing for quicker turnover of results. Positive but low correlations were found between maize seedling and adult root traits, such as number of seminal roots and weight of seminal roots to root pulling resistance (r=0.07 and r=0.36*, respectively) [14]. Expanding the number of seedling root traits and improving respective phenotyping procedures, may increase the chance of capturing strong relationships between different growth stages in maize.

Using digital imaging software to automate phenotypic analysis is an innovative and efficient way of accurately taking measurements of plant physiological traits [15-19]. Roots have been difficult to phenotype in a high throughput manner due to a lack of simple access and their highly plastic nature. With the development of custom root analysis systems, quantitative studies of root systems are now possible [20,21]. There are several software frameworks that extract root morphology traits in two-dimensions in various hierarchies of automation. This ranges from

manual root labeling like DART (Le Bot and Serra, 2009), to semi-automated software like WinRhizo (Pro, 2004), a commercial root analysis tool, and EzRhizo [22], a freely available software, all the way to full integrated imaging-analysis platforms like SmartRoot [23] for small root systems and recent platforms, allowing for automated measurements as well as invoking a ranking system for root traits [17].

These software frameworks have substantially enhanced the research community's ability to efficiently analyze and accumulate massive amounts of data. They also pioneered the utilization of graphical user interface (GUI) that enables ease of use. However, most of these software frameworks are either expensive, not expandable to increased (or decreased) dimensions, or cannot be fine-tuned to a specific setup. We developed an open-source, modular, easy-to-use and efficient root system architecture characterization software called *ARIA* (Automatic Root Image Analysis). This is based on a mathematically rigorous approach of converting root images into graphs. We show how extracting a variety of traits becomes a simple process of utilizing various graph algorithms. There are several major advantages to such a graph based approach to extracting root system traits: (a) graph based methods are well-studied and have very fast and efficient algorithms (for example, used in Google, Facebook, most GPS devices etc.) that enable fast, real time data analysis, (b) graph based methods are easily scalable (having almost linear computational complexity) and, hence, can be easily extended to larger problem sizes without compromising on time (with direct implication to large 3D tomography datasets), and (c) a graph-based approach is generic. That is, by making trivial modifications to the definitions of parameters like edges, weights, and labels, a huge variety of traits can be accessed. This makes a graph based framework trivially extendable. Furthermore, graphs are dimension independent, and hence this framework is trivially extendable to 3D root image analysis.

In this study, the utility of *ARIA* has been tested by phenotyping 384 maize inbred lines using scanned images of seedling roots. These data were then applied to a genome wide association study (GWAS) to detect marker-trait associations. Measurements of the trait Total Root Length were analyzed for a comparative GWAS study, as this is the only trait shared between the current platform WinRhizo Pro 9.0 and ARIA. The objective of this study is to show that our new and freely accessible root phenotyping software *ARIA* is a fast and accurate platform for automated phenotyping, with the potential of adding additional features when compared to the established software WhinRizo Pro 9.0. For both programs, significant marker trait associations were found using a general linear model. Also, phenotypic measurements with both programs were compared using a 74 maize inbred line panel [24] to further validate utility of *ARIA*. The results of this study show that *ARIA* is an accurate and dependable tool for completing large phenotyping experiments, needed for many quantitative genetic studies. Its flexibility makes *ARIA* a very useful tool to breeders and biologists studying root architecture.

Results

Root traits captured by *ARIA*

Using *ARIA*, 27 different root traits were extracted from each scanned image of seedling roots (Table 1). Some traits are more suitable for 3D root scan image analysis such as Depth, Width, and the Width/Depth ratio. All simple statistics as well as heritability estimates for all root traits are found in Table S2. This program is free software and can be accessed using the following link: http://www3.me.iastate.edu/bglab/pages/software.html *ARIA* captures more traits than existing programs such as WinRhizo Pro 9.0, which lists eight different traits that can be obtained from a single root scanned image when buying a standard package. *ARIA* is fully automated with the ability to capture up to three separate seedling roots from a single image, and

to conduct all analyses with limited user interference. Each image was a high resolution scan (around 4400x6200 pixels) of three seedling roots placed side-by-side (Figure 1). Within each image the bounding boxes were automatically identified for each root. Each of the three roots is then individually analyzed and its 27 traits extracted. Data is then exported into an Excel file. This process takes approximately 20 seconds on a standard desktop (2.8 GHz machine). We used a total of 1059 images, each containing up to three roots per image. *ARIA* ran autonomously and extracted traits within 12 hours, allowing for fast turnaround of phenotypic data. Thus, trait capture is very fast and efficient when analyzing multiple roots of large experiments.

Seedling Trait Correlations

Pearson correlations were calculated using SAS 9.3 for all 27 seedling root traits compared to one another. Correlations between traits (Table S1) ranged from very close (r = 0.998) between traits such as secondary root length and PRL to no significant correlations (r = -0.061) for TRL and BSH. BSH did not correlate closely with other root traits with the highest r2 value of 0.166. Similarly, SRL did not show close correlations with other seedling root traits, with its closest correlation of 0.5 with TSA. Conversely, it was found that seedling root trait DEP had close correlations with various other root traits, especially with PRL (r=0.95). A principle component analysis (PCA) was conducted to visualize trait relationships. The first two components explain 45.9% of the variation (with PCA 1 explaining 35.5%). Based on the first two principle components (Figure 2), there are four trait clusters. These clusters are comprised of (1) CMT, WDR, CPT, (2) MNR, and MED, (3) SEL, TRL, NWA, and (4) SCS, WID, PER, CVA, and TSA. All of these traits had close correlations within clusters while traits outside of clusters were not closely correlated (Table S1).

Validation of Measurements

In order to validate measurements made by *ARIA* with those obtained by WinRhizo Pro 9.0 (Regent Instruments, Quebec, Canada), the same images of hydroponically grown maize seedling roots were analyzed by both programs and data compared. Total Root Length was found to be closely correlated with r=0.97 (P=0.0001) when analyzing data within the Ames Panel. For the ASI panel, total root length was correlated between the two programs at r=0.92 (P=0.0001), and root surface area was closely correlated at r=0.90 (P=0.0001). Broad sense heritabilities (H^2) were calculated for both association mapping populations (Table 2). Heritability estimates were generally higher for measurements extracted using *ARIA* at H^2=0.42 compared to H2 = 0.41 for total root length measured in the Ames Panel, as well as root surface area in the ASI panel with H2=0.54 using *ARIA* compared to H^2=0.50 using Whin Rhizo Pro 9.0.

Genome Wide Association Study using *ARIA* vs WinRhizo

A GWAS experiment was conducted in order to show the utility of this new program and its ability to analyze many root images in a high-throughput manner compared to WinRhizo, the current platform used. Further GWAS analyses will be documented in a future publication. TRL was extracted from a single scan of three roots from each inbred line. This process was repeated three times, once for each replication. Analysis of TRL measured with both *ARIA* and WinRhizo combined with genotypic information on 135,311 single nucleotide polymorphism markers across the entire genome identified significant associations at p<5.3 x 10-7. Markers found to be significant were located on chromosomes 1, 2, and 4 for *ARIA* (Figure 3) while WinRhizo analysis resulted in additional SNPs on chromosomes 3, 5, 6, and 8 (Figure 4). Both programs identified significant markers in similar regions of the genome specifically on chromosome 2 and chromosome 4. Moreover, significant SNPs on Chromosome 4 were identical for both programs.

Discussion

Quality of *ARIA* trait estimates, limitations and prospects

ARIA is a reliable program that results in accurate measurements comparable to established programs such as WinRhizo Pro 9.0. The close correlation and higher heritability estimates of TRL and TSA are encouraging for using *ARIA* to obtain accurate measurements in future quantitative studies. A limitation for using *ARIA* in the current study was that only three roots were analyzed at a time. *ARIA* can be extended to allow a larger number of roots to be analyzed within a single image, depending on the scanning or image capture device. Since *ARIA* can automatically crop pictures for the user, keeping roots separate is important for accurate measurements, as crossed over roots could cause uneven cropping or erroneous paths. When comparing the amount of time needed to extract root measurements with *ARIA* and extracting measurements with WinRhizo, *ARIA* simplifies the process and cuts the time taken measurements to less than half the amount of time needed for WhinRhizo. This is in part due to the automatic cropping system as well as exporting measurement values into an Excel spreadsheet all at once within 20 seconds per seedling root. In WhinRhizo, each root has to be cropped manually; data are extracted into a .txt file, which needs to be edited for data analysis. Exporting data directly into a user friendly format *ARIA* by-passes all of these intermediate steps. The current version of *ARIA* is automated for roots exhibiting a distinguishable primary root. However, *ARIA* should work equally well with multiple equal order roots with minor changes due to the graph based formulation: *ARIA* finds all lengths of roots as distances from kernel to root tips and subsequently picks the single longest root (this can be modified to account for multiple equal order roots, for example in rice). A potential limitation is when a secondary root curl ends exactly at the primary root. This creates circular loops in the graph that impair further

analysis. However, none of the 3000 images analyzed exhibited this issue. A way to resolve this minor issue is to consider a quasi-temporal approach to 'growing' the graph vertex-by-vertex that will distinguish these overlaps (work in progress).

The graph based formulation makes this framework easily extendable to multiple purposes [25,26]. This same framework of trait extraction has been applied in other disciplines including chemistry [25] and materials science [26]. Examples of extensions include 3D phenotyping where magnetic resonance imaging (MRI), X-ray or optical tomography data can be curated and traits extracted. Furthermore, *ARIA* can work with a variety of data formats including photographs, scanned images, microscopy images as well as X-ray based reconstructions.

Significance of *ARIA*

While current root analysis programs are available to make measurements of root traits, none currently offers the flexibility and functionality as *ARIA*. When comparing WinRhizo 9.0 to *ARIA*, the larger numbers of traits that can be captured, ability to capture 3D image measurements, and shorter time spent to extract trait measurements from images, are key advantages of. *ARIA*, automatically crops root images, after a mouse click defines the starting point for measurements. Furthermore, *ARIA* has the ability to mark a batch of images enabling batch analysis. In *ARIA*, measurements are exported into an Excel spreadsheet, while WinRhizo gives a text file that must be converted. *ARIA*'s ability to do this automatically makes this program high-throughput and decreases chances of human error. Another key advantage to this program is the fact that measurement capabilities can easily be added, as additional key architectural attributes of roots are determined.

Using *ARIA*, mapping studies for root traits can be implemented on a larger scale due to the reduced time needed for phenotyping. This software system aids plant scientists by relieving the

phenotyping bottleneck for quantitative traits such as root architectural traits by adding to existing technologies in phenomics [27]. Not only is this program fast, its ability to analyze both 2D and 3D images also offers a unique opportunity to look at the same traits, with the same analysis program, but from two different perspectives. Previous programs such as RootReader2D [28] and RootReader3D [29] offer extensive trait collection, but are hindered by the fact that each program is restricted to analyze at either 2D or 3D. *ARIA* in comparison is able to not only analyze 2D flat plane images such as those presented here, but also 3D images of roots. To show this feature, a simple 3D image of a root was analyzed using *ARIA* (Figure 5). Here, we demonstrate that skeletonization and outlining of the primary root can be completed as in 2D. The actual measurements of select traits have also been included in pixels (Figure 5). Based on multiple points of view of the same root system, *ARIA* extracts 27 root traits in a single root analysis. Figure 6 shows how the mathematical foundation (graph based analysis) coupled with the open-source framework can be trivially extended to other trait extraction.

A similar program described by Pascuzzi [17] was used to analyze rice varieties within a gel medium. This program has the ability to capture many of the same traits as *ARIA*. The major advantage of *ARIA* is that it can directly analyze those same gel medium images in both 2D and 3D formats. This adds to the flexibility of this free access program. Existing phenotyping systems can utilize this analysis tool without changing their growth procedures, whereas the other program is not as dimensionally flexible. No changes need to be made in the GUI or procedures to analyze images. This allows for an expanded number of environmental conditions, whether controlled by humans or nature, in which root architecture could be studied and for connections between how root develop in a hydroponic environment compared to soil or other growth medium.

Exploring Roots as a model for selection

Large scale mapping studies such as quantitative trait locus (QTL) mapping and GWAS require large mapping populations that must be phenotyped in an accurate manner. Genomic selection (GS) [30] is a method in which a training population is used to collect phenotype information and coupled with extensive genetic information. Then, a model is developed to make predictions for the performance of traits of interest, solely based on genetic information. This requires massive amounts of phenotypic information that are highly accurate, especially for plastic traits such as root architecture [31]. Currently, root architecture is not used for selection, because of the resources needed for extensive phenotyping and the quantitative mode of inheritance of root traits [9]. New phenotyping software such as *ARIA* may facilitate to include root architecture in selection schemes. Comparative GWAS for TRL obtained with both WinRhizo Pro 9.0 and *ARIA* identified similar or identical regions of the genome associated with this trait. Associations found in only one program may be due to low power of detecting a polymorphism with small genetic effect.

The major goal was to develop an easy to use image software analysis tool for measuring root traits from simple scans or photographs. A free to use software platform with ability to investigate both 2D and 3D root architectural characteristics for plants has been developed to facilitate measuring multiple root traits in a high-throughput, accurate manner. We compared this new program to existing programs. *ARIA* showed close correlations to traits measured with established software, supporting accurate measurements. The 27 root traits measured give an example for the utility of this program and offer an extensive amount of traits to be studied for large scale phenotypic analysis of roots or mapping studies looking at the genetic control of root architecture. Future studies using this program include root characterization for particular maize

or other plant species of interest as well as phenotyping for quantitative trait studies such as GWAS, QTL mapping, and GS.

Materials and Methods

***ARIA* (Automatic Root Image Analysis)**

ARIA is custom software written in the programming framework, MATLAB (Natick, Massachusetts, United States). *ARIA* has a user friendly GUI interface to enable easy and rapid data extraction. The operational concept of the software is to convert the root image (after standard image pre-processing) into a graph. The software framework can read in most standard image formats. Each image is loaded (Figure 1), and after a sequence of pre-processing steps, converted into a graph. A graph is a mathematical construct consisting of a set of vertices that are connected by a set of edges. This is done by labeling each pixel of the root image into a vertex, and linking nearest neighbor pixels with edges. The key steps of the software are:

a) **Thresholding**: The background is first identified (using morphological operations in Matlab) and renormalized to black. This effectively eliminates most of the background signal. Then the image threshold is calculated using Otsu's method. The grey scale image is converted into a black and white image. This is done by comparing the intensity of each pixel with a threshold value. The pixel is marked as black (or white), if it's grey scale value is smaller (or larger) than the threshold (Figure 7).

b) **Connected components**: Since the root is one large connected system, everything else that is not connected to the root can be removed from the image. This idea is encoded in the graph concept of connected components, which enumerates all the distinct connected components in the image. The largest connected component is the root, all the other connect components are

noise or other foreign artifacts. Note that if the image resolves finer root hairs (which our imaging process does not do) these will still be part of the largest connected component.

c) **Skeletonization:** A 'wire-frame' skeleton of the binary image is constructed by thinning (or eroding). Skeletonization is a fundamental tool with many applications in image processing and visualization. Here, skeletonization is essential to identify and distinguish between the primary and secondary roots (Figure 7).

d) **Primary and secondary root identification:** The primary root is identified as the graph path that has the longest path length (Figure 8). This is accomplished by Dijkstra's algorithm to estimate shortest paths between two points of the graph [32]. Dijkstra's algorithm is used to compute the shortest paths from each free end of the root to every other free end. The longest "shortest path" is identified as the primary root. Secondary roots are identified easily by subtracting the primary root from the original image and enumerating the remaining distinct connected components.

e) **Graph querying and post processing:** The graph is queried to construct several traits starting from simple traits like total root length, to more complex measures like bushiness. All data are exported into an Excel sheet for ease of analysis and use. This will allow one to place a series of images for analysis at a time and export it to Excel. The data are also displayed on the GUI. All traits are analyzed automatically and can be viewed when clicking display results (Figure 9).

Plant materials

The first association mapping population or "Ames panel" is comprised of 384 inbred lines obtained from the USDA-ARS North Central Regional Plant Introduction Station (NCRPIS) in Ames, Iowa. All lines used in this study are a subset of a larger collection of lines called the Ames panel [33], consisting of 2815 maize inbred lines conserved at USDA-ARS NCRPIS. The

384 lines were selected based on maturity in view of future field trials in central Iowa. The second panel of 74 maize inbred lines called "ASI panel" includes ex-PVPs (Plant Variety Protection) and Germplasm enhancement of Maize (GEM) inbred lines [24].

Root phenotyping

Cigar Roll Growth Conditions

A paper roll assay described by [24] was used for germination and growth of maize seedlings. Seedlings were grown in 2 L glass beakers filled with 1.4 L of sterilized water. Seedlings were placed in a growth chamber for 14 days at 16/8 hrs light/darkness (25/22 ^{0}C). Light intensity was 200 µmol photons m-2s-1, and a relative humidity maintained at 65%. Each paper roll with four seedlings was considered as experimental unit. After 14 days seedlings were removed from the growth chamber and phenotypic traits measured. If not all traits were measured the same day, plants were preserved in 30% ethanol to prevent aging of roots.

Image Acquisition

Seedling roots were imaged using a high resolution scanner. Three separate seedling roots were imaged at a time using an EPSON Expression 10000 XL scanner system (Copyright © 2000-2014 Epson America, Inc).

Phenotype Data Analysis

Experimental Design

Ames panel lines were grown in three experiments starting June 12, 2012, July 3, 2012, and October 5, 2012. Each experiment was grown in the same growth chamber and at the same growing conditions, as described above. Lines were grown in a completely randomized design (CRD) and trait data were collected per experimental unit: three seedlings out of four within each seed roll were sampled, to eliminate possible outliers within lines, and means taken. The ASI

panel of 74 maize inbred lines were grown under the same conditions and replicated twice under one experiment. Analysis of variance of root traits was performed, the additive model for analysis of variance was:

$$yij = \mu + R_i + G_j + E_{ij}$$

where y_{ij} represents the observation from the ijth experimental unit, μ is the overall mean, R_i is the ith experiment and G_j is the jth genotype. The interaction between the fixed effects G_j and the random effect experiment is confounded with the error E_{ij}. The statistics software package SAS 9.3 (Copyright © 2014 SAS Institute Inc.) was used to obtain ANOVA tables, expected mean squares, and least square means for association analyses. Function PROC GLM was implemented and type 3 sums of squares were used to account for missing data. Genotypic (σ_g^2), and phenotypic (σ_p^2) variances as well as broad sense heritability (H^2) were all calculated on an entry mean basis. Heritability on an experimental unit basis was calculated as follows:

$$H^2 = \frac{\sigma_G^2}{\sigma_P^2},\ \sigma_G^2 = \left(\frac{MSG - MSE}{rep\ \#}\right),\ \sigma_P^2 = \left(\frac{MSG - MSE}{rep\ \#}\right) + MSE,\ H^2 = \frac{\left(\frac{MSG - MSE}{rep\ \#}\right)}{\left(\frac{MSG - MSE}{rep\ \#}\right) + MSE}$$

Function PROC GLM was implemented. Pearson correlations were calculated using the SAS function CORR to determine the relationships between seedling traits.

Marker Data

Genotyping-by-sequencing (GBS) [34] was used to genotype the association mapping population with 681,257 single nucleotide polymorphism (SNP) markers across the maize genome. Imputation as described by [33] was employed. In an effort to reduce the number of non-informative markers, all monomorphic SNP markers and those with more than 20% missing data were omitted. SNP markers with a minor allele frequency less than 5% were removed, leaving

135,311 SNP markers spread across all 10 chromosomes of the maize genome to calculate population structure, kinship, and to perform GWAS.

Association analyses

Population structure was estimated from a reduced number of unimputed SNPs (1,665 SNP markers) using program Structure 2.3.4 [35]. Parameter settings for estimating membership of coefficients of coancestry for lines are a burn-in length of 50,000 with 50,000 iterations for each cluster (K) from 1-15, with each K being run five times. We applied an admixture model with independent allele frequencies. To pick the most probable K value, we used an ad hoc (ΔK) statistic based on the ordering rate of change of P(X|K) [36]. Software program TASSEL 4.0 [37] was used to calculate LD as well as Loiselle kinship coefficients between lines based on 135,311 SNP markers. Population structure (Q matrix) was used in association analyses to decrease the amount of type 1 errors [38]. TASSEL 4.0 was used to conduct genome wide association analyses (GWAS) using a General Linear Model (GLM) and population structure as a fixed factor with model $y = X\beta + U$, where y are the values measured, X is the marker value, β is a matrix of parameters to be estimated, and U uses the Q values as fixed factors. To account for multiple testing during GWAS, statistical package simpleM was implemented in R 3.0 [39]. Based on a α level of P=0.05, the multiple testing threshold level was set to 5.3x10-7 with the equation α/n , where n equals the effective number of independent tests. Only the Ames panel was analyzed, as genomic marker data were not available for the ASI panel.

References

1. Aiken RM, Smucker AJM (1996) ROOT SYSTEM REGULATION OF WHOLE PLANT GROWTH. Annual Review of Phytopathology 34: 325-346.

2. Lynch J (1995) Root Architecture and Plant Productivity. Plant Physiology 109: 7-13.

3. Feldman L (1994) The Maize Root. In: Freeling M, Walbot V, editors. The Maize Handbook: Springer New York. pp. 29-37.

4. Hochholdinger F, editor (2009) Handbook of Maize: Its Biology. 145 p.

5. Hoppe DC, McCully ME, Wenzel CL (1986) The nodal roots of Zea: their development in relation to structural features of the stem. Canadian Journal of Botany 64: 2524-2537.

6. Zhu J, Mickelson SM, Kaeppler SM, Lynch JP (2006) Detection of quantitative trait loci for seminal root traits in maize (Zea mays L.) seedlings grown under differential phosphorus levels. Theor Appl Genet 113: 1-10.

7. Liu J, Li J, Chen F, Zhang F, Ren T, et al. (2008) Mapping QTLs for root traits under different nitrate levels at the seedling stage in maize (Zea mays L.). Plant and Soil 305: 253-265.

8. Jordan WR, Douglas WA, Shouse PJ (1983) Strategies for Crop Improvement for Drought-Prone Regions. Agriculture Water Management 7: 281-299.

9. Salvi RTaS (2007) From QTLs to Genes controlling Root Traits in Maize. pp. 13-22.

10. Trachsel S, Kaeppler SM, Brown KM, Lynch JP (2010) Shovelomics: high throughput phenotyping of maize (Zea mays L.) root architecture in the field. Plant and Soil 341: 75-87.

11. Hammer GL, Dong Z, McLean G, Doherty A, Messina C, et al. (2009) Can Changes in Canopy and/or Root System Architecture Explain Historical Maize Yield Trends in the U.S. Corn Belt? Crop Science 49: 299.

12. Landi P (1998) Seedling characteristics in hydroponic culture and field performance of maize genotypes with different resistance to root lodging. Maydica 43: 111.

13. Landi P, Giuliani M, Darrah L, Tuberosa R, Conti S, et al. (2001) Variability for root and shoot traits in a maize population grown in hydroponics and in the field and their relationships with vertical root pulling resistance. Maydica 46: 177-182.

14. Nass HG, Zuber MS (1971) Correlation of Corn (Zea mays L.) Roots Early in Development to Mature Root Development1. Crop Sci 11: 655-658.

15. Brewer MT, Lang L, Fujimura K, Dujmovic N, Gray S, et al. (2006) Development of a Controlled Vocabulary and Software Application to Analyze Fruit Shape Variation in Tomato and Other Plant Species. Plant Physiology 141: 15-25.

16. Chavarria-Krauser A, Nagel KA, Palme K, Schurr U, Walter A, et al. (2008) Spatio-temporal quantification of differential growth processes in root growth zones based on a novel combination of image sequence processing and refined concepts describing curvature production. New Phytol 177: 811-821.

17. Iyer-Pascuzzi AS, Symonova O, Mileyko Y, Hao Y, Belcher H, et al. (2010) Imaging and analysis platform for automatic phenotyping and trait ranking of plant root systems. Plant Physiol 152: 1148-1157.

18. Nagel KA, Putz A, Gilmer F, Heinz K, Fischbach A, et al. (2012) GROWSCREEN-Rhizo is a novel phenotyping robot enabling simultaneous measurements of root and shoot growth for plants grown in soil-filled rhizotrons. Functional Plant Biology 39: 891.

19. Wang L, Uilecan IV, Assadi AH, Kozmik CA, Spalding EP (2009) HYPOTrace: Image Analysis Software for Measuring Hypocotyl Growth and Shape Demonstrated on Arabidopsis Seedlings Undergoing Photomorphogenesis. Plant Physiology 149: 1632-1637.

20. Le Bot J, Serra V, Fabre J, Draye X, Adamowicz S, et al. (2009) DART: a software to analyse root system architecture and development from captured images. Plant and Soil 326: 261-273.

21. Zeng G, Birchfield ST, Wells CE (2008) Automatic discrimination of fine roots in minirhizotron images. New Phytol 177: 549-557.

22. Armengaud P, Zambaux K, Hills A, Sulpice R, Pattison RJ, et al. (2009) EZ-Rhizo: integrated software for the fast and accurate measurement of root system architecture. The Plant Journal 57: 945-956.

23. Lobet G, Pagès L, Draye X (2011) A novel image-analysis toolbox enabling quantitative analysis of root system architecture. Plant physiology 157: 29-39.

24. Abdel-Ghani AH, Kumar B, Reyes-Matamoros J, Gonzalez-Portilla PJ, Jansen C, et al. (2012) Genotypic variation and relationships between seedling and adult plant traits in maize (Zea mays L.) inbred lines grown under contrasting nitrogen levels. Euphytica 189: 123-133.

25. Wodo O, Roehling JD, Moulé AJ, Ganapathysubramanian B (2013) Quantifying organic solar cell morphology: a computational study of three-dimensional maps. Energy & Environmental Science 6: 3060.

26. Samudrala S, Wodo O, Suram SK, Broderick S, Rajan K, et al. (2013) A graph-theoretic approach for characterization of precipitates from atom probe tomography data. Computational Materials Science 77: 335-342.

27. Furbank RT, Tester M (2011) Phenomics--technologies to relieve the phenotyping bottleneck. Trends Plant Sci 16: 635-644.

28. Clark RT, Famoso AN, Zhao K, Shaff JE, Craft EJ, et al. (2013) High-throughput two-dimensional root system phenotyping platform facilitates genetic analysis of root growth and development. Plant Cell Environ 36: 454-466.

29. Clark RT, MacCurdy RB, Jung JK, Shaff JE, McCouch SR, et al. (2011) Three-dimensional root phenotyping with a novel imaging and software platform. Plant Physiol 156: 455-465.

30. Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of Total Genetic Value Using Genome-Wide Dense Maker Maps. Genetics 157: 1819-1829.

31. Gruber BD, Giehl RF, Friedel S, von Wiren N (2013) Plasticity of the Arabidopsis root system under nutrient deficiencies. Plant Physiol 163: 161-179.

32. Knuth DE (1977) A generalization of Dijkstra's algorithm. Information Processing Letters 6: 1-5.

33. Romay MC, Millard MJ, Glaubitz JC, Peiffer JA, Swarts KL, et al. (2013) Comprehensive genotyping of the USA national maize inbred seed bank. Genome Biol 14: R55.

34. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, et al. (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6: e19379.

35. Pritchard JK, Stephens M, Donnelly P (2000) Inference of Population Structure Using Multilocus Genotype Data. Genetics 945-959.

36. Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14: 2611-2620.

37. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, et al. (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23: 2633-2635.

38. Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, et al. (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38: 203-208.

39. Gao X, Starmer J, Martin ER (2008) A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. Genet Epidemiol 32: 361-369.

Table I. ***Traits captured by ARIA***

Trait Name	Symbol	Trait Description
Total Root Length	TRL	Cumulative length of all the roots in centimeters
Primary Root Length	PRL	Length of the Primary root in centimeters
Secondary Root Length	SEL	Cumulative length of all secondary roots in centimeters
Center of Mass	COM	Center of gravity of the root.
Center of Point	COP	Absolute center of the root regardless of root length.
Center of Mass (Top)	CMT	Center of gravity of the top 1/3 of the root (Top).
Center of Mass (Mid)	CMM	Center of gravity of the middle 1/3 root (Middle).
Center of Mass (Bottom)	CMB	Center of gravity of the bottom 1/3 root (Bottom).
Center of Point (Top)	CPT	Absolute center of the root regardless of root length (Top).
Center of Point (Mid)	CPM	Absolute center of the root regardless of root length (Middle).
Center of Point (Bottom)	CPB	Absolute center of the root regardless of root length (Bottom).
Maximum Number of Roots	MNR	The 84th percentile value of the sum of every row.
Perimeter	PER	Total number of network pixels connected to a background pixel.
Depth	DEP	The maximum vertical distance reached by the root system.
Width	WID	The maximum horizontal width of the whole RSA.
Width/Depth ratio	WDR	The ratio of the maximum width to depth.
Median	MED	The median number of roots at all Y-location.
Total Number of Roots	TNR	Total number of roots.
Convex Area	CVA	The area of the convex hull that encloses the entire root image
Network Area	NWA	The number of pixels that are connected in the skeletonized image
Solidity	SOL	The fraction equal to the network area divided by the convex area
Bushiness	BSH	The ratio of the maximum to the median number of roots.
Length Distribution	LED	The ratio of TRL in the upper one-third of the root to the TRL.
Diameter	DIA	Diameter of the primary root.
Volume	VOL	Volume of the primary root
Surface Area	SUA	Surface area of the primary root.
SRL	SRL	Total root length divided by root system volume

Table II. ***Comparison of repeatability estimates for both WhinRhizo Pro 9.0 and ARIA***

Analyzing tool	Trait	Heritability (H^2)
WinRhizo Pro 2009	Total Root Length (Ames Panel)	.41
ARIA	Total Root Length (Ames Panel)	.42
WinRhizo Pro 2009	Total Root Length (ASI Panel)	.42
ARIA	Total Root Length (ASI Panel)	.42
WinRhizo Pro 2009	Root Surface area (ASI Panel)	.50
ARIA	Root Surface area (ASI Panel)	.54

Supplementary Table S1. Trait correlations between all 28 traits extracted using *ARIA*. Non-significant correlations denoted with '*'

TRL	SUA	PRL	SEL	COM	COP	CMT	CMM	CMB	CPT	CPM	CPB	MNR	PER	DEP	WID	WDR	MED	MRN	CVA	NWA	SOL	LED	DIA	VOL	SCS	SRL	BSH
1.000	0.873	0.696	0.981	0.205	-0.022*	-0.09	0.277	-0.171	-0.06	0.114	-0.242	0.397	0.733	0.684	0.655	-0.03*	0.891	0.884	0.818	0.983	.087**	-0.24	0.244	0.379	0.662	-0.364	-0.061*
	1.000	0.768	0.872	0.170	-0.10	-0.121	0.254	-.149	-0.070	0.101	-0.216	0.434	0.743	0.733	0.658	-0.103	0.760	0.745	0.820	0.885	-0.059*	-.232	0.393	0.465	0.799	0.500	-0.065
		1.000	0.670	0.017*	-0.308.	-0.229	0.268	-0.206	-0.16	0.092	-0.248	0.635	0.821	0.954	0.671	-0.174	0.459	0.427	0.864	0.712	0.321	-0.085	0.154	0.396	0.859	-0.285	-0.099*
			1.000	0.219	0.006*	-0.082	0.276	-0.145	-0.057*	0.116	-0.219	0.373	0.721	0.659	0.654	-0.075	0.925	0.919	0.811	0.998	0.042	-0.236*	0.255	0.387	0.681	-0.389	0.065*
				1.000	0.368	0.275	0.229	-0.043*	0.097	0.118	-0.127	0.149	0.198	-0.036*	0.120	0.046*	0.298	0.264	0.151	0.208	0.153	-0.499	0.258	0.121	0.121	-0.248	0.026*
					1.000	0.282	0.178	-0.011*	0.262	0.330	-0.021*	-0.13	-0.162	-0.363	-0.096	0.146	0.158	0.166	-0.153	-0.019*	0.364	-0.345	0.275	-0.007*	-0.137	-0.249	-0.032*
						1.000	-0.027*	0.010*	0.550	0.023	0.006*	-0.058*	-0.077	-0.243	-0.124	-0.012*	-0.021	0.028*	-0.16	-0.096	0.068*	-0.138	0.098	-0.075	-0.144	0.074*	-0.019*
							1.000	-0.268	0.017	0.752	-0.278	0.271	0.306	0.269	0.235	0.017*	0.262	0.278	0.287	0.282	0.051*	-0.154	0.176	0.135	0.259	-0.283	0.009*
								1.000	-0.002*	-0.214	0.904	-0.305	-0.255	-0.213	-0.155	-0.105	-0.113	-0.139	-0.183	-0.153	-0.096	0.115	-0.125	-0.053*	-0.179	0.187	-0.166
									1.000	0.047*	-0.007	0.114	-0.022*	-0.169	-0.113	-0.036*	-0.013*	0.044*	-0.136	-0.066	0.041*	-0.093	0.098	-0.057*	-0.094	-0.104	-0.058*
										1.000	-0.222	0.141	0.126	0.104	0.091	0.051*	0.134	0.169	0.108	0.116	0.112	-0.08	0.119	0.005*	0.077*	-0.227	0.003*
											1.000	-0.356	-0.304	-0.245	-0.196	-0.094	-0.185	-0.204	-0.244	-0.226	-0.101	0.163	-0.123	-0.066*	-0.204	0.196	-0.202
												1.000	0.698	0.604	0.374	-0.125	0.215	0.252	0.521	0.401	-0.19	-0.168	0.213	0.288	0.579	-0.289	-.044*
													1.000	0.785	0.695	-0.118	0.574	0.575	0.849	0.746	-0.210	-0.183	0.241	0.390	0.760	-0.369	-0.076
														1.000	0.582	-0.214	0.428	0.403	0.811	0.698	-0.317	-0.018*	0.049*	0.311	0.771	-0.199	-0.110
															1.000	-0.001*	0.564	0.564	0.880	0.669	-0.29	-0.050*	0.313	0.406	0.687	-0.370	-0.080
																1.000	-0.039*	-0.031*	-0.079*	-0.021	0.755	-0.043*	0.044*	-0.016*	-0.104	-0.072	0.038*
																	1.000	0.936	0.656	0.911	0.193	-0.326	0.278	0.302	0.486	-0.392	-0.066*
																		1.000	0.632	0.902	0.210	-0.210	0.309	0.291	0.462	-0.426	-0.058*
																			1.000	0.632	-0.245	-0.147	0.278	0.451	0.821	-0.385	-0.079
																				1.000	0.045*	-0.236	0.251	0.386	0.678	-0.388	-0.065*
																					1.000	-0.211	0.075	-0.09	-0.226	-0.122	0.072
																						1.000	-0.364	-0.233	-0.229	0.300	-0.004*
																							1.000	0.450	0.587	-0.808	0.104
																								1.000	0.499	-0.407	-.005*
																									1.000	-0.594	-0.068*
																										1.000	-0.032*
																											1.000

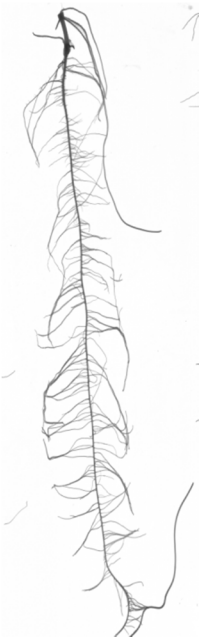


Figure 1: Image of a 14 day root

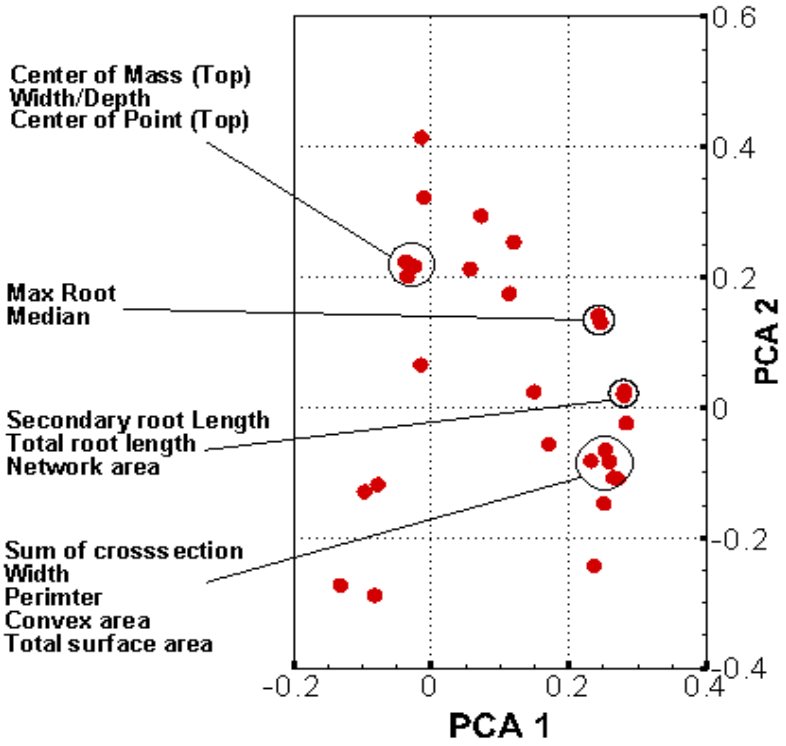


Figure 2. PCA plot of all *ARIA* traits Pearson correlations, clusters of traits have been marked showing traits are closest related.

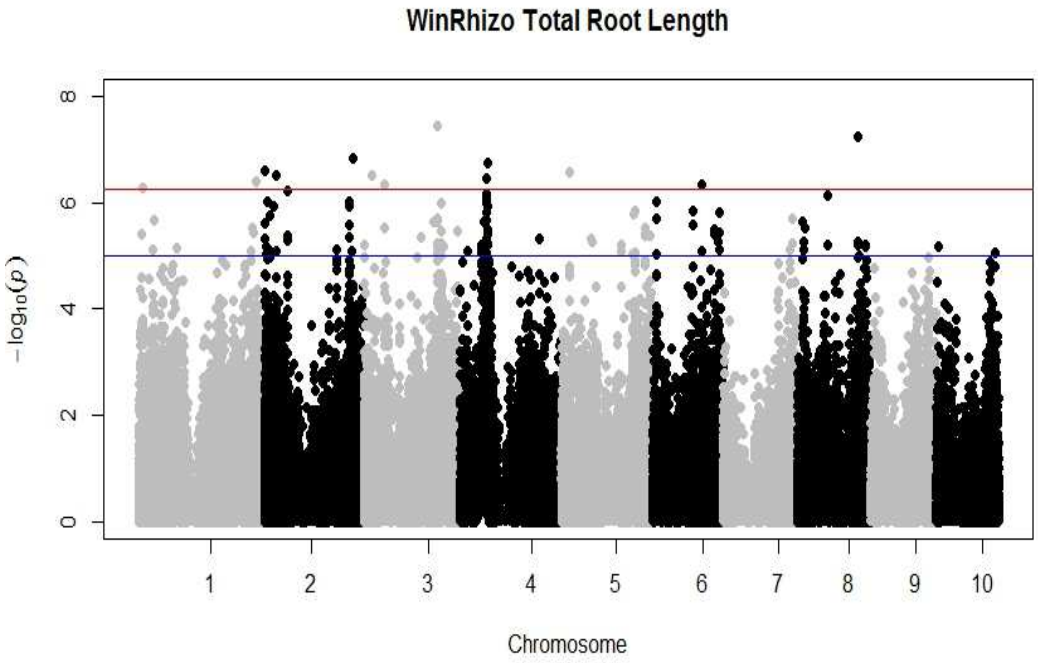


Figure 4. Manhattan plot displaying all 10 maize chromosomes, showing SNP markers significantly associated with trait Total Root Length measured with WinRhizo, significant SNPs are consistent with *ARIA* with additional SNPs on chromosomes 3, 5, 6, and 8.

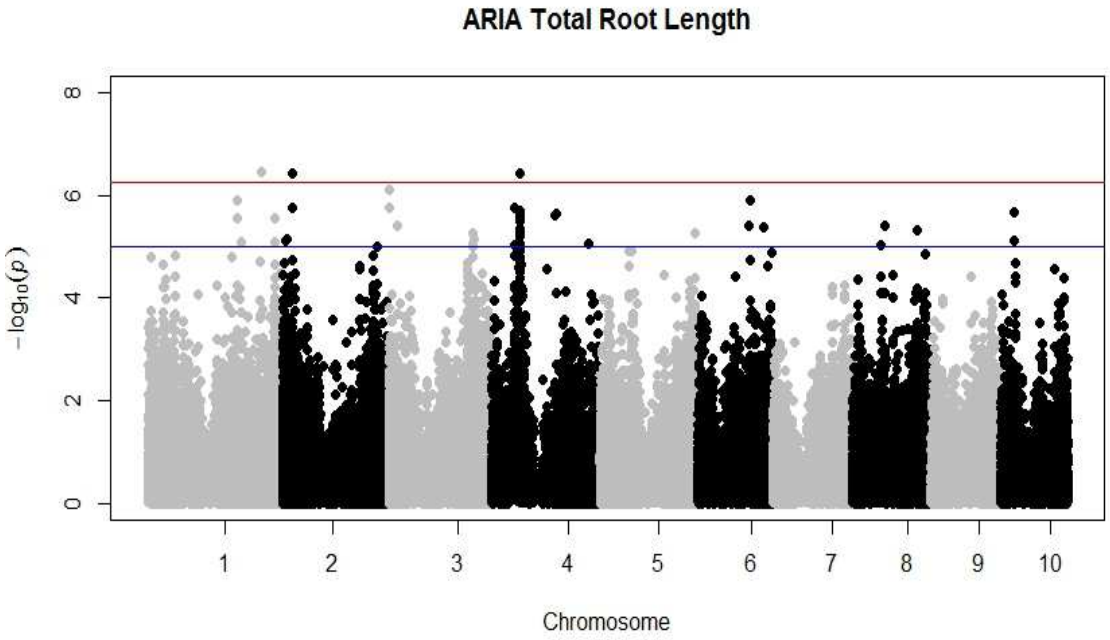


Figure 3. Manhattan plot displaying all 10 maize chromosomes, showing SNP markers significantly associated with trait Total Root Length measured with *ARIA*. Significant SNPs are located on chromosomes 1, 2, and 4.

Figure 6. Analysis of a diseased maple leaf, *ARIA's* flexible framework will allow multiple uses of the program beyond root phenotyping. While we have not fully explored this capability, *ARIA* will likely also be useful to extract above ground traits such as leaf vein structure, and disease quantification.

Figure 7: Thresholding and skelotinization stages

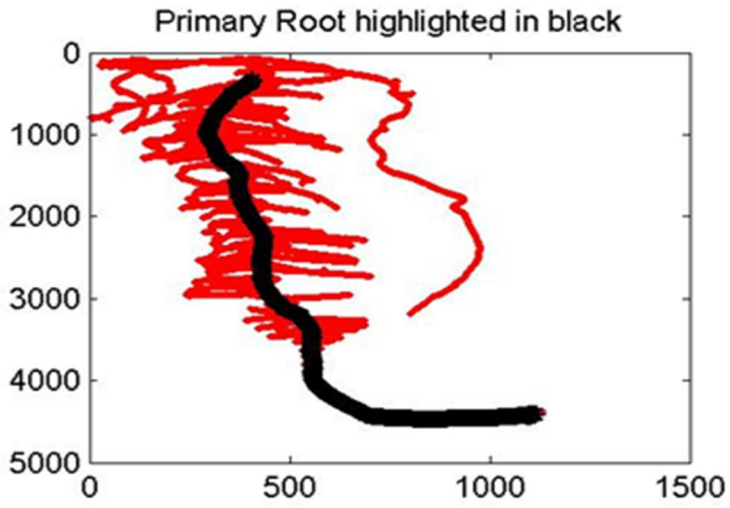


Figure 8: Automated identification of primary and secondary roots

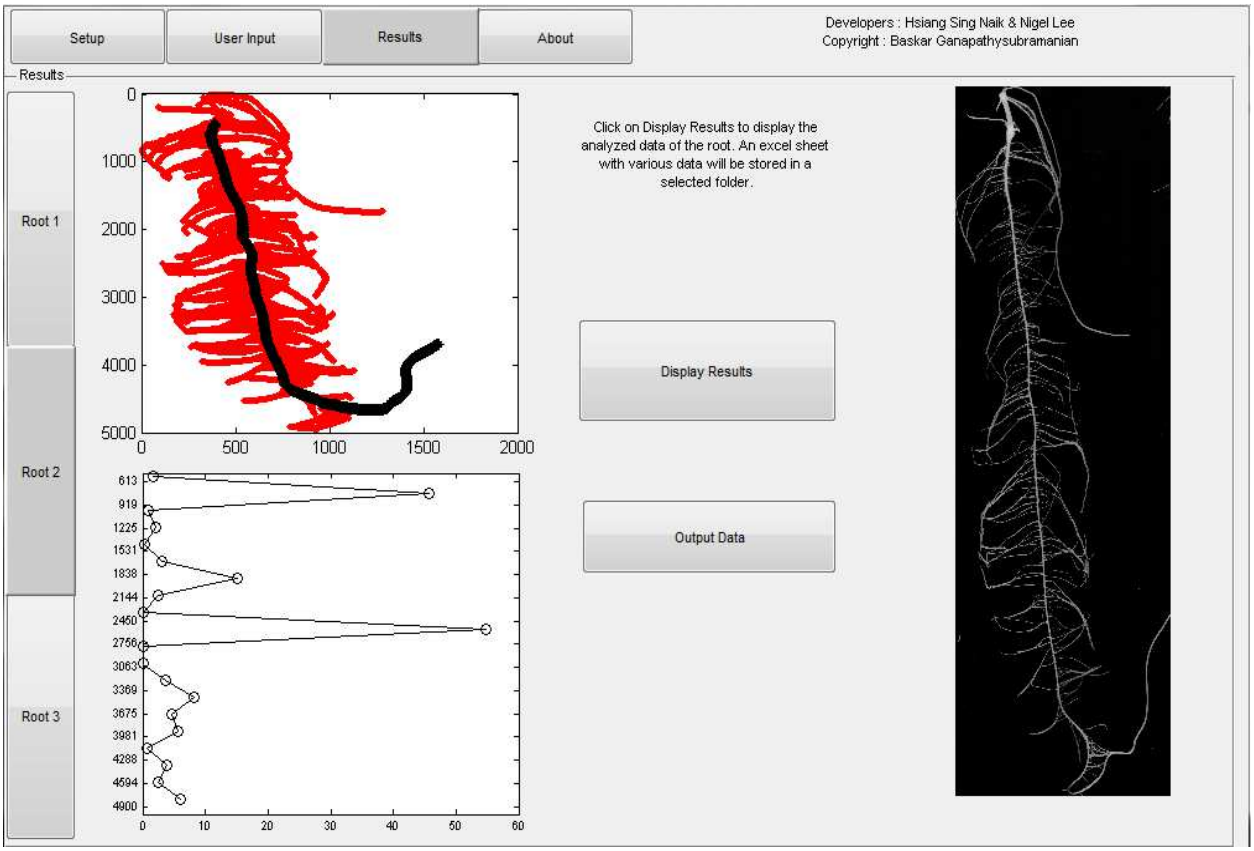


Figure 9: Screen capture of the *ARIA* framework. The picture on the right is the root image. The plot on the top left is automated identification of primary and secondary roots. The graph on the bottom left represents detailed analysis of root architecture, specifically a histogram of secondary roots across each 10% of the primary root.

CHAPTER FOUR

GENOME-WIDE ASSOCIATION ANALYISIS OF SEEDLING ROOT DEVELOPMENT IN MAIZE (*ZEA MAYS* L)

Jordon Pace, Candice Gardner, Baskar Ganapathysubramanian, Thomas Lübberstedt

Paper published in BMC Genomics Journal. Abstract, structure, and references are formatted according to the journal standards.

Abstract

Background: Plants rely on the root system for anchorage to the ground and the acquisition and absorption of nutrients critical to sustaining productivity. A genome wide association analysis enables one to analyze allelic diversity of complex traits and identify superior alleles. 384 inbred lines from the Ames panel were genotyped with 681,257 single nucleotide polymorphism markers using Genotyping-by-Sequencing technology and 22 seedling root architecture traits were phenotyped.

Results: Utilizing both a general linear model and mixed linear model, a GWAS study was conducted identifying 268 marker trait associations ($p \leq 5.3x10^{-7}$). Analysis of significant SNP markers for multiple traits showed that several were located within gene models with some SNP markers localized within regions with previously identified root quantitative trait loci. Gene model GRMZM2G153722 located on chromosome 4 contained nine significant markers. This predicted gene is expressed in roots and shoots.

Conclusion: This study identifies putatively associated SNP markers associated with root traits at the seedling stage. Some SNPs were located within or near (< 1 kb) gene models. These gene models identify possible candidate genes involved in root development at the seedling stage.

These and respective linked or functional markers could be targets for breeders for marker assisted selection of seedling root traits.

Background

In an effort to increase crop production, farmers and producers apply millions of tons of fertilizers such as Nitrogen (N) each year. In 2010, demand for N fertilizer was 103.9 million tons and is expected to steadily increase to 111 million tons by 2014 worldwide [1]. Only around 33% of the N applied is taken up by cereal crops such as maize [2],[3], while the remaining N is lost due to a combination of factors including leaching, de-nitrification, and surface runoff from the soil. These issues affect the environment and input costs negatively [2],[4].

The root system is essential for plant species to absorb and acquire mineral nutrients such as N. Plant species such as maize (Zea mays L.) have two general mechanisms to increase nutrient acquisition: 1) develop a larger root system that allows plants to come into contact with a larger soil volume, and 2) increase the trans-membrane nutrient-uptake rate. Increased root size allows plants to increase available nutrient uptake based on demand within a limited time frame [5]. Root architecture and development has been shown to be a key component in nitrogen use efficiency (NUE) [6], and drought tolerance [7]. Understanding root development and the molecular mechanisms that influence root architecture is thus important for increasing yield potential and yield stability under varying environmental conditions and soil profiles [8].

Maize has five main types of roots: crown, seminal, primary, lateral, and brace roots [9]. The primary and seminal roots make up the embryonic root system and their fate is largely determined by genetic background [9]. The major portion of adult root biomass is derived from postembryonic shoot-borne roots, crown roots which are formed below the soil surface and brace

roots which are formed above the soil surface [10]. Lateral roots are initiated from the pericycle of other roots and have a strong influence on maize root architecture [11]. Their function is important to plant performance as they are responsible for a crucial part of water and nutrient uptake, such as N in maize. It has been shown that N rich soil environments enhance root growth and dry weight [12]. Root size has been shown to be a key component in the uptake of phosphorus, calcium, in addition to N [12],[13]. Increasing root size and, therefore, root surface area might be a strategy plants use to increase absorption efficiency, when nutrients such as N are limiting [14]. Thus genomic regions affecting root development and growth could affect NUE, water use efficiency, and nutrient use efficiency as roots with increased root length and surface area may perform better in nutrient deficient environments. Several genes have been described that affect the development of the root system in maize including Rtcs (rootless, concerning crown and seminal roots), Rth1 (roothairless1), Rth3 (roothairless 3), and Rum1 (rootless with undetectable meristems1). Rtcs controls crown root and seminal root formation; Rtcn and Rtcl are thought to be paralogs of Rtcs. Rth1 and Rth3 control root hair elongation with Rth3 being shown to affect grain yield in maize [15],[16]. While these genes have been identified, there are many loci effecting root growth and development that remain unknown.

A useful method for analyzing the genetic diversity of complex traits and identification of superior alleles is association mapping or linkage disequilibrium (LD) mapping [17]. Unlike traditional linkage mapping, where bi-parental populations are developed, association mapping uses ancestral recombination in natural populations to find marker-phenotype associations based on LD [18]. Association mapping allows evaluation of a large number of alleles in diverse populations [19], and offers additional advantages compared to traditional linkage mapping, including high mapping resolution and reduction in time to develop a mapping population [20].

There are two main association mapping strategies. The candidate gene approach focuses on polymorphisms in specific genes controlling traits of interest, while genome-wide association approaches survey the entire genome for polymorphisms associated with complex traits [21]. A candidate gene association analysis approach was employed using genes Rtcl, Rth3, Rum1, and Rul1[22]. Several polymorphisms within all four candidate genes were associated with seedling root traits. Many of these significant polymorphisms affected putative functional sequence motifs including transcription factor binding sites and major domains. Another study [23] used 73 elite Chinese maize lines to investigate sequence variation and haplotype diversity for the root development gene Rtcs. They too found extensive variation between lines at the gene sequence level. The advent of more economic sequencing technologies facilitates genome-wide studies. Using markers covering the entire genome increases the chance of identifying additional regions of the genome associated with seedling root traits, and establishing relevance of above mentioned candidate genes to other genes affecting root development. In this study, a panel of 384 inbred lines derived from the Ames panel [24] was used to conduct a genome-wide association study (GWAS) to investigate root architecture at the seedling stage. Our hypothesis is that root architecture is of quantitative inheritance and that there are multiple factors throughout the genome that contribute to root development. The objectives of this study were to i) study phenotypic variation of 22 root architecture traits within a maize association panel, ii) identify SNP markers throughout the genome associated with root architecture traits, and iii) investigate locations of associated SNP markers for possible candidate genes or functional markers having an effect on root development.

Results

Analysis of Phenotypes of 384 Ames Panel Inbred Lines

Almost all root traits captured followed a normal distribution with a slight left skew. Trait descriptions are found in Table 1 and Additional file 1: Figure S1. Most traits had considerable variation within the current mapping population. The standard deviation for traits such as Total Root Length (TRL) and Secondary Root Length (SEL) varied the most with values of 98.07 and 92.8 respectively. All trait maximum, minimum, and standard deviations are listed in Table 2. A few lines' phenotypes were consistently placed in the tails of the distribution for multiple traits. Line PHT77 had the highest values for TRL, SEL, Surface Area (SUA), and Network Area (NWA). These traits are all highly and significantly ($P < 0.0001$) correlated with one another (Table 2) with $r = 0.90$. NWA is also highly correlated with root Median (MED) and Total Number of Roots (TNR), yet PHT77 doesn't have the highest values for these traits. This can be due to many reasons, one being that much of PHT77's root length comes from the individual length of its secondary roots; this also increases root Surface Area (SUA) and NWA. This also lowers PHT77's TNR and MED as there are fewer number of secondary roots present for this maize line. A243 showed the lowest values for root Perimeter (PER), TNR, MED, and Maximum Number of Roots (MNR). Interestingly, these traits were significantly ($P < 0.0001$) but not always, closely correlated, ranging from $r = 0.27$ to 0.95. Heritability (H2) estimates for all traits were low to moderate and ranged from 0.12 to 0.49 (Table 2). Due to the low heritability estimates of some traits, and in accordance with other similar studies analyzing root traits [19], a cutoff of $H^2 \geq 0.30$ was made, and most traits with $H2 < 0.30$ were excluded from further analysis.

Pearson correlations were calculated comparing the same traits (TRL and total plant biomass) (TBP) measured in a previous association panel [25] that used the same measuring techniques as in this study by comparing lines that were the same between both mapping populations. This was done to determine, if growing conditions were consistent and if ARIA calculated measurements were consistent with result obtained from image analysis software WhinRHIZO Pro 9.0. Both traits were significantly correlated ($p = 0.05$) between both methods with values of $r = 0.85$ for TRL and $r = 0.75$ for TPB (data not shown).

Correlation coefficients were calculated for the 22 traits listed in Table 3. The two traits with the closest correlation were TRL and SEL ($r = 0.98$), indicating that much of the root system is made up of lateral and seminal root length, not the primary root at the 14 day old seedling stage. Correlations were lower between TRL and Primary Root Length (PRL) ($r = 0.72$) and between PRL and SEL ($r = 0.68$). Correlations for 1000 kernel seed weight (KRW) were also calculated to determine whether kernel size had a major effect on seedling root size, which was collected prior to growing plants in the growth chamber. None of the seedling traits collected showed a strong ($r = 0.33$) correlation with kernel weight (data not shown).

Linkage Disequilibrium Decay in Ames Panel Subset

A random subset of markers spanning across all 10 chromosomes (see Methods) was used to calculate LD decay. The rate of LD decay was similar across chromosomes with an average distance of reaching the LD threshold ($r^2=0.2$) within approximately 10 kb throughout the genome. Chromosome 8 showed the slowest decay with an r^2 value of 0.2 reached at approximately 15 kb (Figure 1). These results are comparable to [24], indicating that LD decayed within 1-10 kb.

Population Structure

In order to define the number of subpopulations within the 384 line Ames panel subset, the ad hoc statistic (ΔK) was calculated. Based on the ad hoc statistic values in Structure 2.3.4 the mapping population was sorted into two subpopulation (K=2). One subpopulation comprised of 319 lines or 83% of the total 384 lines used for GWAS (Figure 2). This larger subpopulation is composed of mostly non-stiff stalk inbred lines with some tropical, popcorn, and mixed lines. The other subpopulation includes mostly genotypes from the stiff-stalk heterotic group. B73 is found within this subpopulation whereas Mo17 is found in the larger subpopulation.

Genome-Wide Association Studies

Four SNP markers were found to be significantly associated with two root traits using MLM. The threshold to account for multiple testing was determined by simpleM at $P = 5.36 \times 10^{-7}$. Specifically, one significant marker-trait association was found for Bushiness (BSH) located on chromosome 2 (Figure 3), and three significant SNP marker trait associations for Standard Root Length (SRL) were located on chromosome 3 (Figure 4). Based on heritability estimates both traits were found below the threshold to be examined in depth. Due to the stringency of MLM, and the fact that significant markers found for both traits are located in regions of the genome consistent with significant markers for other root traits using GLM, it was decided that these significant SNPs be used for further examination. All three significant markers for SRL were found within gene models. Marker S4_49565840 was found within gene model GRMZM2G327349, expression analysis based on B73 showed very little to no expression within roots. The two other markers (S4_49619564 and S4_49619525) significantly associated with SRL were found within gene model GRMZM2G32186. This gene model did show expression

both at germination and at V1 stage of maize development in the primary root with absolute expression levels of 7385.82 and 5539.36 respectively (Table 4). The one significant marker for BSH on chromosome 2 was found within gene model GRMZM2G322186 and showed very little to no expression in the roots throughout early development. No other traits were found to have significant marker trait associations using the Q + K MLM model.

Using the GLM model, an additional 287 significant markers were found using the same threshold of $P=5.36 \times 10^{-7}$ for root traits above the heritability threshold of $H^2 \geq 0.30$. Clustering of significant SNPs using GLM was analyzed. SNPs associated with root traits clustered on chromosomes 2, 3, 4, and 8 (B73 reference genome 2). Chromosome 2 also contained the SNP marker with the highest significance. Most significant markers on chromosome 2 were located in bins 2.00–2.02 and 2.07-2.08. Clusters on chromosome 3 were located within bins 3.01 and 3.06-3.09 while clusters on chromosome 4 were within bin 4.05. On chromosomes 2 and 8, four markers in total were significantly associated with multiple traits. Chromosome 2 had 3 markers; marker S2_20263530 was significant for PRL, PER, Diameter (DIA), Depth (DEP), Shoot Dry Weight (SDW), TBP, and SUA. Marker S2_202178253 on chromosome 2 is associated with traits SUA, SDW, SL, and TPB. The third and final marker on chromosome 2 was marker S2_20252886; this marker is associated with both SUA and TBP. These three significant markers are found within gene models GRMZM2G002879, GRMZM2G154864, and GRMZM2G087254. The final marker is S8_146152722 and was associated with both PER and DEP. This marker on chromosome 8 is located in gene model GRMZM2G070837. On chromosome 4, 13 markers were found significantly associated with multiple traits. All 13 markers on chromosome 4 are located within 250 kb. Nine of these markers are located within the same gene model, GRMZM2G153722. Of the remaining four markers on chromosome 4,

two are located in the same gene model GRMZMZG427409; one is located in another gene model GRMZM2G053511 while the remaining marker is located in an intergenic region. Four of the previously listed gene models have hypothetical protein products. An earlier expression analysis [26] revealed that most of the predicted gene models described above had moderate to low expression levels in the primary root system at growth stage V1 in B73. Absolute expression levels measured in B73 for respective gene models are listed in Table 3. When looking at SNPs close to previously reported genes with an impact on root development (Rtcs, Rtcl, Rul1, Rum1, and Rth1), one significant SNP marker at position 205,392,941 on chromosome 3 is located a little more than 3 Mbs from Rum1. No other significant markers were located in or near previously reported root development genes. A list of all significant marker trait associations is found in Additional file 2: Table S1. Manhattan plots for all marker trait associations using GLM are found in Additional file 3: Figure S2.

Discussion

Root traits are difficult and laborious to measure at the adult stage in a field setting. In the current study, measurements of seedling root architectural traits in our association mapping population were used as a first step for later comparison with adult plant traits. One of the traits studied, RDW, has been shown to be positively correlated with key adult plant traits such as yield at both HN and LN conditions [25], suggesting that seedling root traits may be useful to predict adult root characteristics. One concern with studying seedling roots is that seed size might be confounded with overall seedling vigor including expression of root traits. However, all seedling root traits had low correlations (r-values <0.33) with kernel weight.

Root architecture is a key plant characteristic but highly variable among maize genotypes. Table 1 demonstrates this wide range of variation for most traits studied herein. For TRL, a 9- to10-fold difference was found within the current mapping population, specifically three lines (Va38, NO. 1201 INBRED, and INBRED 309) that were all recorded as having the lowest TRL average measurements and the three lines with the longest average root length (PHT77, Mo1W, and PHK29). This range exceeded the 3- to 4-fold differences in a separate, albeit smaller (72 lines) association panel [25]. This large range for average length of roots illustrates the extensive amount of phenotypic variation found for roots. This range in trait values among inbred lines can be compared to other studies of diverse maize panels [27], where there was a 3- and 2-fold difference for plant height and days to anthesis, respectively. In conclusion, there is substantial unexploited variation for root traits.

Heritability values ranged from 0.12 to 0.49. Previous studies have shown similar ranges of heritabilies for root traits at various stages of growth, both under controlled environmental (growth chamber, greenhouse) and field conditions [19],[28]. Root growth is highly plastic and of quantitative nature. By keeping all conditions equal, some root traits were more repeatable than others. Biomass traits (TPB, RDW, SDW, and Shoot Length (SHL)) as well as TNR had mid-range heritabilities close to 0.5. Other traits that deal with total length of roots or a particular part of the root (TRL and SEL) also had heritabilities greater than 0.4. This may be due to the software ARIA's ability to accurately measure length based traits. Some traits with low heritabilities in our study of 2D traits may be better suited for three dimensional images such as BSH, DEP, Length Distribution (LED), and Width/Depth ratio (WDR). PRL showed a low heritability estimate (H^2=0.281). This could be due to limitations in ARIA's ability to identify the primary root accurately each time, or is a product of PRL sensitivity to micro environmental

conditions. We included PRL in the present study, as this trait has been shown to be important in water and nutrient acquisition [11].

Population Structure and Linkage Disequilibrium

Population structure analysis using the software package Structure 2.3.4 [29] revealed two subpopulations. The two identified populations fit the two major heterotic groups within temperate U.S. maize germplasm: stiff stalk (with B73) and non-stiff stalk (including Mo17). The larger subpopulation contained over 82% of the lines in the association panel, this subpopulation was made up of non-stiff stalk inbred and few mixed heterotic group lines. These results are consistent with results from a principle component analysis (PCoA) of the entire Ames Panel consisting of over 2800 lines [24]. In that study, most lines derived from the U.S. grouped in two distinct groups, stiff stalk and non-stiff stalk.

Average LD decay (r^2 threshold = 0.2) across the whole genome was close to 10 kb. These results agree with a LD decay of 10 kb across ExPVP, stiff stalk, and non-stiff stalk lines within the entire 2,815 inbred lines within Ames Panel [24]. Romay et al 2013, used the same GBS marker data set in order to analyze the entire Ames Panel diversity. The subset of inbred lines from the Ames panel used in this study lacks diversity from tropical lines that are available within the complete Ames panel. If more exotic maize germplasm is included as in other association mapping populations, the rate of decay is usually more rapid (around 300 bp-1 kb) with added diversity [24, 30].

Association Analysis

There have been several large scale genome-wide association studies which have been used to identify candidate genes and putative functional markers that affect complex traits [19, 31-33]. In

the current study, four SNPs were significantly associated with root traits BSH, and SRL using the Q+K MLM. When fitting just population structure using GLM, 287 SNPs were significantly associated with root traits. Among those, 17 were significantly associated with multiple root traits. Three of these 17 SNPs were located in similar positions on chromosome 2. SNP S2_202635930 was significantly associated with seven traits, PRL, PER, DIA, DEP, SDW, TPB, and SUA. All seven traits are closely and significantly correlated with one another ($r>0.5$). This trend continued for all traits sharing significant SNPs: all were significantly correlated with one another (Table 3). Other SNPs associated with multiple traits were located on chromosomes 4 and 8. Three root QTL studies [28, 34, 35] identified a QTL on chromosome 4 within bin 4.05-4.07. In this region, 13 of the shared, significantly associated SNPs were located. These results provide evidence that relevant candidate genes affecting root growth and development are likely located on chromosome 4.

The only two traits (SRL and BSH) for which significant SNPs were detected using MLM had low heritability estimates. Since associations were found fitting both the Q and K matrix, the risk of type I error is low. BSH and SRL are components of other traits (Table 1). Thus, significant polymorphisms for BSH and SRL might act pleiotropic and affect traits with higher heritability. For a few traits, no significantly associated markers were detected (Width (WID), Convex Area (CVA), SEL, and Center of Point (COP)). The number of detected associations was not related to heritability. TPB had the highest heritability estimate with $H^2 = 0.491$ and 17 significant SNPs were detected for this trait, while only two SNPs were detected for TNR with comparably high heritability (0.49). Conversely, 135 SNPs markers were significantly associated with Diameter (DIA) ($H^2=0.33$). Different reasons may account for this discrepancy, such as (i) tight linkage of multiple associated SNPs for a low heritability trait, (ii) absence of detectable SNPs in genome

regions impacting high heritability traits, and (iii) unknown trait architecture, i.e., number of genes and distribution of gene effects with impact on traits of interest.

GLM is less stringent than MLM. This explains the large discrepancy between vastly different numbers of significant associations detected by the two methods of calculation. As noted in other studies [36], MLM can over fit a model and create type II errors. Thus, using both methods in conjunction is preferable. We made an effort to reduce type I error using GLM by fitting the Q matrix, and by applying correction for multiple testing. Even though only few significant polymorphisms were identified using MLM, those were co-located in clusters of significantly associated polymorphisms identified by using GLM.

Candidate Genes for seedling root traits

For the MLM analysis, gene model GRMZM26322186 contained two of the significant markers for seedling root trait SRL. This candidate gene is expressed throughout seedling development [26]. It should be noted that these expression information is based on B73, and variation in transcriptome profiles between multiple inbred lines has been reported [37]. The gene model codes for three putative protein products within maize *Zea* CEFD homolog1, TPA: isopenicillin N epimerase isoform 2 and isoform 1. No confirmed function of these proteins has been determined.

The most noticeable candidate gene identified within this study is GRMZM2G153722. Its gene model is located on chromosome 4 and contained 12 of the 13 significant markers found for two traits, DIA and SUA. Haplotype analysis for this gene was examined with two haplotypes being identified within this region of the genome. One haplotype was found significant for both DIA and SUA at p-values of 5.22×10^{-9} and 2.66×10^{-8} respectively. This strengthens our findings at

the individual SNP level. Throughout seedling development this gene model showed expression is detectable in both roots and shoots [26]. The candidate gene is predicted to code for a putative protein 1-phospatidylinositol-4-phosphate 5-kinase. A BLAST search identified homologues in two species, *Sorghum bicolor* and *Setaria italica* (foxtail millet), with greater than 85% sequence identity. Both species have hypothetical protein products with currently unknown function. A homologue in Arabidopsis thaliana [38] plays an important role in root tip growth. If the function of the respective maize gene is similar, this candidate gene could be a vital player in regulating root development.

Gene models GRMZM2G154864 and GRMZM2G322186 contained significant SNPs for multiple traits. BLAST results for GRMZM2G154864 cDNA identified both *Sorghum bicolor*, bamboo, and *Setaria italica* with greater than 85% sequence identity, as was previously noted for GARMZM2G153722. Results from a BLAST of GRMZM2G322186 cDNA revealed 100% identity with maize gene *cef1*, which codes for an aspartate aminotransferase (AAT) superfamily (fold type I) gene of pyridoxal phosphate (PLP)-dependent enzymes. No phenotypes have been linked to this putative gene and protein product. Expression of these genes was detected at V1 stage in the primary root in B73 [26]. These genes could play an active role in root development, especially at seedling growth stage.

Wild type alleles of root development genes *Rtcl, Rth3, Rum1*, and *Rul1* were studied with regard to their impact on seedling root trait expression using a candidate gene-based association mapping approach [39]. SNPs within these genes among the 72 inbred lines used as a mapping population were found to be associated with both root traits RDW and TRL. In our study, *Rum1* was putatively detected by a linked significant SNP. No SNPs within the remaining genic regions were significantly associated in this study. We used a candidate gene-based association

mapping approach for those same four candidate genes in our population to determine, whether any SNP in these regions would show significance due to a less stringent multiple testing threshold. Nevertheless, we did not find significant SNPs within these root development genes. Most lines used for the previous association panel [39] were different from those in our panel, which would affect the significance of SNPs within those specific genes. In the previous study, Sanger sequencing was used, which resulted in almost complete information of polymorphic sites within above mentioned candidate genes, giving much finer resolution within these specific genic regions. The same polymorphisms were likely not included within the current imputed GBS data due to different alleles being found in the different populations [37]. These differences in allele frequencies could lead to more or fewer loci being polymorphic within these genic regions. For example in the previous study for root gene *Rtcl*, 45 polymorphisms were detected. In our population only five SNPs were present within this region. Due to these discrepancies in allelic frequencies between populations, it can be expected that results can be inconsistent between association studies in different panels [24].

Conclusions

The putative SNPs identified within the current study might aid in selecting lines with these particular phenotypic root characteristics. Respective SNPs can be used to breed for specific root types under various environmental conditions, thus enabling use of maize root architecture information as part of a selection strategy. The idea of an ideal root architecture or root ideotype has been presented [40-43]. Ideotypes such as “steep, deep, and cheap” roots [40], or deeper roots with vigorous lateral root growth, may increase nitrogen uptake efficiency under low N conditions [42]. Other root traits that might play a pivotal role in increasing nutrient uptake efficiency include seminal root length and number, lateral root length and number, and root

distribution. Due to the extensive resource requirements needed to study adult plant roots, being able to connect seedling root traits to adult plant traits would be beneficial. Understanding consistent relationships between seedling and adult root architectural traits would enable selection at the seedling level, and is addressed in ongoing research.

Methods

Plant materials

The association mapping panel consists of 384 inbred lines obtained from the USDA-ARS North Central Regional Plant Introduction Station (NCRPIS) in Ames, Iowa (Supplementary Table 2). These 384 lines are a subset of the "Ames panel" [24], a collection of 2,815 maize inbred lines conserved at the USDA-ARS NCRPIS. The 384 lines were selected on the basis of maturity in view of future evaluations in central Iowa, genetic diversity, and geographic origin, with preference for dry climates that might require vigorous root development. Thirteen lines from a previous experiment [25] were duplicated in our association panel including B73, Mo17, and PHZ51.

Root Phenotyping

Paper Roll Experiments

A paper roll growth method was employed as described by [22]. Briefly, seed was sterilized using Clorox® solution (6% sodium hypochlorite) for 15 minutes. After soaking, seed was twice rinsed in autoclaved water. Brown germination roll paper (Anchor Paper, St. Paul, MN, USA) was pre-moisturized with fungicide solution Captan® (2.5g/l) before being vertically rolled, with four kernels per genotype and growth paper roll. Germination paper rolls were placed in two liter

glass beakers containing 1.4 liters of autoclaved deionized water, at a photoperiod of 16/8 hrs (light/darkness) and 25/22 ^{0}C. Light intensity was 200 µmol photons m-2s-1, and a relative humidity maintained at approx. 65%. Each paper roll with four seedlings was considered an experimental unit. After 14 days, seedlings were removed from the growth chamber and all root traits were measured. If not measured the same day, plants were preserved in 30% ethanol to prevent aging of roots.

Manually evaluated traits were root dry weight (RDW), shoot dry weight (SDW), shoot length (SHL), and total plant biomass (TPB). SHL was measured manually using a ruler measuring from the base of the shoot to the tip of the primary leaf. After root and shoot measurements were conducted, roots and shoots were collected separately and dried for 48 hrs at 55^{0}C, to determine RDW, SDW, and TPB. In addition, 22 traits (Table 1 and Supplementary Figure 1) were determined using *ARIA* (Automatic Root Image Analyzer) high-throughput phenotyping software [44]. For this purpose, roots of each genotype were placed on a scanner to produce high resolution images.

Phenotypic Data Analysis

Experimental design

Our association panel was grown in a completely random design (CRD) in three independent replications completed June 13, 2012, July 3, 2012, and October 5, 2012. Each experiment was grown in the same growth chamber under the same growing conditions. All trait data for phenotypic analysis were collected on a plot basis (plot is equal to our experimental unit: three seedlings out of four within each seed roll were sampled, to eliminate possible outliers within lines, and means were taken). The additive model for analysis of variance was:

$$y_{ij} = \mu + R_i + G_j + E_{ij.}$$

Where y_{ij} represents the observation from the ij^{th} plot, μ is the overall mean, R_i is the experiment and G_j is the genotype. The interaction between the fixed effect G_j and the random effect experiment is confounded with the error ($E_{(ij)}$). The statistical software package SAS 9.3 was used to obtain ANOVA table, expected mean squares, and least square means for association analyses. Function PROC GLM was implemented and type 3 sums of squares were used to account for missing data. Genotypic (σ_g^2), and phenotypic (σ_p^2) variances as well as broad sense heritability (H^2) (due to the fact that we cannot partition out additive variance alone) were all calculated on an entry mean basis. Heritability was calculated as follows:

$$H^2 = \frac{\sigma_G^2}{\sigma_P^2}, \quad \sigma_G^2 = (\frac{MSG - MSE}{rep}), \quad \sigma_P^2 = (\frac{MSG - MSE}{rep}) + MSE, \quad H^2 = \frac{(\frac{MSG - MSE}{rep})}{(\frac{MSG - MSE}{rep}) + MSE}$$

MSG and MSE stand for mean square of genotype and mean square error, respectively. Rep is the number of independent replications (3). Least square means across all three replications were calculated using SAS 9.3 to adjust means. Phenotypic correlations were calculated using the SAS function PROC CORR to determine the relationship between seedling traits.

Marker data

Genotyping-by-sequencing (GBS) [45], was used to genotype all inbred lines with 681,257 markers distributed across the entire maize genome. GBS uses the restriction enzyme ApeKI and is run on an Illumina platform. The current data set was obtained using 96 sample multiplexes per Illumina flow cell. A total of 681,257 bilallelic SNP markers were distributed across all 10 chromosomes of the maize genome, imputation was used to reduce the number of missing data

points. The imputation algorithm uses a nearest neighbor approach based on 64 base SNP windows across the entire maize sequence database allowing for 5% mismatches [24]. Biallelic markers with a minor allele frequency below 10% were removed from the marker data set. All monomorphic SNP markers and those with more than 20% missing data were omitted. Finally, 135,311 SNP markers distributed across all 10 chromosomes of the maize genome with a slight bias towards telomeric regions remained to calculate population structure, kinship, and to perform GWAS.

Population Structure, Linkage Disequilibrium, and Association Analysis

Population structure (Q matrix) was estimated from a reduced random number of unimputed SNPs (1,665 SNP markers) using Structure 2.3.4 software [29]. The parameter settings for estimating membership coefficients for lines in each subpopulation were a burn-in length of 50,000 followed by 50,000 iterations for each of the clusters (K) from 1-15, with each K being run five times. An admixture model was applied with independent allele frequencies. In order to pick the most probable K value for analysis, a method [46] calculating an ad hoc (ΔK) statistic based on the ordering rate of change of $P(X|K)$ was employed.

The software program *TASSEL* 4.0 [47] was used to calculate LD and to conduct GWAS using a General Linear Model (GLM) using population structure as a fixed factor with an equation of $y = X\beta + U$, where y equals the values measured, X is the marker value, β is a matrix of parameters to be estimated, and U uses the Q values as fixed cofactors to account for errors and false positives caused by population substructure. LD decay, or the distance in base pairs that loci could be expected to be in LD, was calculated by plotting r^2 onto genetic distance in measured in base pairs using an r^2 value of 0.2 as a cutoff. All markers with less than 35% missing data and a

minor allele frequency greater than 0.05% were used to calculate LD decay. Once r^2 values were calculated, this data was summarized using R 3.0 statistical software for each of the 10 maize chromosomes individually as well as combining all chromosomes to test a genome wide LD decay. Software SpAGeDi [48] was used to calculate the Loiselle kinship coefficients between lines (K matrix).

A mixed linear model (MLM) was also used for association studies utilizing the program GAPIT (Genome Association and Prediction Integrated Tool-R package) [49]. Statistical model for MLM was $y = X\beta + Zu + e$. Terms X, and Z are incidence matrices of 1s and 0s, X relates β to term y and Z relates u to y. The term y is a vector of the phenotypic values. Term β is an unknown fixed effect that represents marker effects and population structure (Q), u is a vector of size n (n representing the number of individuals, 384 for this population) for random polygenetic effects having a distribution with mean of zero and covariance matrix of $G = 2K\sigma^2G$. Where K is the kinship matrix, used to determine correlations between different individuals and determine whether they are independent, as our assumption is that all individuals are independent from one another. Both Q and K matrices were fit in the MLM to control spurious associations due to population structure and relatedness, respectively [50]

To account for multiple testing during GWAS, the statistical program simpleM was implemented using R program 3.0 [51]. Based on an α level of 0.05, the multiple testing threshold level was set at 5.3x10-7. This threshold is based on an effective number of independent tests of n, Meff_G. To obtain Meff_G for SNP data, a correlation matrix for all markers needs to be constructed and corresponding eigenvalues for each SNP locus calculated. A composite LD (CLD) correlation is calculated directly from SNP genotypes [49]. Once this SNP matrix is created, the effective number of independent tests is calculated and this number is used in a

similar way as the Bonferroni correction method. Here, the alpha level threshold ($\alpha = 0.05$) is divided by Meff_G (α/(Meff_G)). Markers above the suggested threshold for MLM were considered as significantly trait-associated SNP markers and candidates for causative polymorphisms.

References

1. FAO: **Current world fertilizer trends and outlook to 2014**. In. Rome: Food and Agriculture Organization of the United Nations; 2010.
2. Tilman D, KGC, Matson P, Naylor R, and Polasky S: **Agriculture sustainability and intensive production practices**. *Nature* 2002, **418**.
3. Raun WR, Johnson GV: **Improving Nitrogen Use Efficiency for Cereal Production**. *Agronomy Journal* 1999, **91**(3):357-363.
4. Quemada LM: **Strategies to Improve Nitrogen Use Efficiency in Winter Cereal Crops under Rainfed Conditions**. *Agron J* 2008, **100**(2):277-284.
5. Tian Q, Chen F, Zhang F, Mi G: **Genotypic Difference in Nitrogen Acquisition Ability in Maize Plants Is Related to the Coordination of Leaf and Root Growth**. *Journal of Plant Nutrition* 2006, **29**(2):317-330.
6. Hirel B, Le Gouis J, Ney B, Gallais A: **The challenge of improving nitrogen use efficiency in crop plants: towards a more central role for genetic variability and quantitative genetics within integrated approaches**. *J Exp Bot* 2007, **58**(9):2369-2387.
7. Ribaut J-M, Fracheboud Y, Monneveux P, Banziger M, Vargas M, Jiang C: **Quantitative trait loci for yield and correlated traits under high and low soil nitrogen conditions in tropical maize**. *Molecular Breeding* 2007, **20**(1):15-29.

8. Hodge A, Berta G, Doussan C, Merchan F, Crespi M: **Plant root growth, architecture and function**. *Plant and Soil* 2009, **321**(1-2):153-187.

9. Hochholdinger F: **The Maize Root System: Morphology, Anatomy, and Genetics.** In *Handbook of Maize: Its Biology*; 2009. Edited by Bennetzen JL and Hake SC.

10. Hoppe DC, McCully ME, Wenzel CL: **The nodal roots of Zea: their development in relation to structural features of the stem**. *Canadian Journal of Botany* 1986, **64**(11):2524-2537.

11. Lynch J: **Root Architecture and Plant Productivity**. *Plant physiology* 1995, **109**:7-13.

12. Mackay S: **Root Growth and phosphorus and potassium uptake by two corn genotypes in the field**. *Fertilizer Research* 1986, **10**:217-230.

13. Marschener H: **Role of root growth, arbuscular mycorrhiza, and root exudates for the efficiency in nutrient acquisition**. *Field Crops Research* 1998, **56**(1-2):203-207.

14. Horst WJ, Kamh M, Jibrin JM, Chude VO: **Agronomic measures for increasing P availability to crops**. *Plant and Soil* 2001, **237**:211-223.

15. Taramino G, Sauer M, Stauffer JL, Multani D, Niu X, Sakai H, Hochholdinger F: **The maize (Zea mays L.) RTCS gene encodes a LOB domain protein that is a key regulator of embryonic seminal and post-embryonic shoot-borne root initiation**. *Plant J* 2007, **50**(4):649-659.

16. Hochholdinger F, Wen TJ, Zimmermann R, Chimot-Marolle P, da Costa e Silva O, Bruce W, Lamkey KR, Wienand U, Schnable PS: **The maize (Zea mays L.) roothairless 3 gene encodes a putative GPI-anchored, monocot-specific, COBRA-like protein that significantly affects grain yield**. *Plant J* 2008, **54**(5):888-898.

17. Yan J, Warburton M, Crouch J: **Association Mapping for Enhancing Maize (L.) Genetic Improvement**. *Crop Science* 2011, **51**(2):433.

18. Buckler ES, Thornsberry JM: **Plant molecular diversity and applications to genomics**. *Curr Opin Plant Biol* 2002, **5**:107-111.

19. Krill AM, Kirst M, Kochian LV, Buckler ES, Hoekenga OA: **Association and linkage analysis of aluminum tolerance genes in maize**. *PloS one* 2010, **5**(4):e9958.

20. Yu J, Buckler ES: **Genetic association mapping and genome organization of maize**. *Curr Opin Biotechnol* 2006, **17**(2):155-160.

21. Risch N, Merkangas K: **The Future of Genetic Studies of Complex Human Diseases**. *Science* 1996, **273**.

22. Narayana BKT: **Candidate gene based association study for nitrogen use efficiency and associated traits in maize**. *Dissertation* 2013:pgs 74-102. Iowa State University. http://lib.dr.iastate.edu/etd/13017/.

23. Zhang E, Yang Z, Wang Y, Hu Y, Song X, Xu C: **Nucleotide Polymorphisms and Haplotype Diversity of RTCS Gene in China Elite Maize Inbred Lines**. *PloS one* 2013, **8**(2).

24. Romay MC, Millard MJ, Glaubitz JC, Peiffer JA, Swarts KL, Casstevens TM, Elshire RJ, Acharya CB, Mitchell SE, Flint-Garcia SA *et al*: **Comprehensive genotyping of the USA national maize inbred seed bank**. *Genome Biol* 2013, **14**(6):R55.

25. Abdel-Ghani AH, Kumar B, Reyes-Matamoros J, Gonzalez-Portilla PJ, Jansen C, Martin JPS, Lee M, Lübberstedt T: **Genotypic variation and relationships between seedling and adult plant traits in maize (Zea mays L.) inbred lines grown under contrasting nitrogen levels**. *Euphytica* 2012, **189**(1):123-133.

26. Sekhon RS, Lin H, Childs KL, Hansey CN, Buell CR, de Leon N, Kaeppler SM: **Genome-wide atlas of transcription during maize development**. *Plant J* 2011, **66**(4):553-563.

27. Peiffer JA, Romay MC, Gore MA, Flint-Garcia SA, Zhang Z, Millard MJ, Gardner C, McMulen MD, Hollan JB, Bradbury PJ, Buckler ES: **The Genetic Architecture of Maize Height**. *Genetics* 2014, **196**(4):1337-1356.

28. Cai H, Chen F, Mi G, Zhang F, Maurer HP, Liu W, Reif JC, Yuan L: **Mapping QTLs for root system architecture of maize (Zea mays L.) in the field at different developmental stages**. *TAG Theoretical and applied genetics Theoretische und angewandte Genetik* 2012, **125**(6):1313-1324.

29. Pritchard JK, Stephens M, Donnelly P: **Inference of Population Structure Using Multilocus Genotype Data**. *Genetics* 2000, **155**:945-959.

30. Wu X, Li Y, Shi Y, Song Y, Wang T, Huang Y, Li Y: **Fine genetic characterization of elite maize germplasm using high-throughput SNP genotyping**. *TAG Theoretical and applied genetics Theoretische und angewandte Genetik* 2014, **127**(3):621-631.

31. Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, Ersoz E, Flint-Garcia S, Garcia A, Glaubitz JC *et al*: **The genetic architecture of maize flowering time**. *Science* 2009, **325**(5941):714-718.

32. Brown PJ, Upadyayula N, Mahone GS, Tian F, Bradbury PJ, Myles S, Holland JB, Flint-Garcia S, McMullen MD, Buckler ES *et al*: **Distinct genetic architectures for male and female inflorescence traits of maize**. *PLoS Genet* 2011, **7**(11):e1002383.

33. Weng J, Xie C, Hao Z, Wang J, Liu C, Li M, Zhang D, Bai L, Zhang S, Li X: **Genome-wide association study identifies candidate genes that affect plant height in Chinese elite maize (Zea mays L.) inbred lines**. *PloS one* 2011, **6**(12):e29229.

34. Zhu J, Kaeppler SM, Lynch JP: **Mapping of QTL controlling root hair length in maize (Zea mays L.) under phosphorus deficiency**. *Plant and Soil* 2005, **270**(1):299-310.

35. Zhu J, Kaeppler SM, Lynch JP: **Mapping of QTLs for lateral root branching and length in maize (Zea mays L.) under differential phosphorus supply**. *TAG Theoretical and applied genetics Theoretische und angewandte Genetik* 2005, **111**(4):688-695.

36. Xue Y, Warburton ML, Sawkins M, Zhang X, Setter T, Xu Y, Grudloyma P, Gethi J, Ribaut JM, Li W *et al*: **Genome-wide association analysis for nine agronomic traits in maize under well-watered and water-stressed conditions**. *TAG Theoretical and applied genetics Theoretische und angewandte Genetik* 2013, **126**(10):2587-2596.

37. Hansey CN, Vaillancourt B, Sekhon RS, Nata, De Leon N, Kaeppler SM, Buell CR: **Maize (Zea mays L.) Genome Diversity as Revealed by RNA-Sequencing**. *PloS One Genetics* 2012, **7**(3).

38. Kusano H, Testerink C, Vermeer JEM, Tsuge T, Shimada H, Oka A, Munnik T, Aoyama T: **The *Arabidopsis* Phosphatidylinositol Phosphate 5-Kinase PIP5K3 Is a Key Regulator of Root Hair Tip Growth.** *Plant Cell* 2008, **20**: 367-380.

39. Zhang Z, Ersoz E, Lai CQ, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu J, Arnett DK, Ordovas JM *et al*: **Mixed linear model approach adapted for genome-wide association studies**. *Nat Genet* 2010, **42**(4):355-360.

40. Lynch JP: **Steep, cheap and deep: an ideotype to optimize water and N acquisition by maize root systems**. *Annals of botany* 2013, **112**(2):347-357.

41. Donald CM: **The breeding of crop ideotypes**. *Euphytica* 1968, **17**(3):385-403.

42. Mi G, Chen F, Wu Q, Lai N, Yuan L, Zhang F: **Ideotype root architecture for efficient nitrogen acquisition by maize in intensive cropping systems**. *Sci China Life Sci* 2010, **53**(12):1369-1373.

43. Postma JA, Dathe A, Lynch J: **The optimal lateral root branching density for maize depends on nitrogen and phosphorus availability**. *Plant physiology* 2014.

44. Pace J, Lee N. Naik HS, Ganapathysubramanian B, Lubberstedt T: **Analysis of maize (Zea mays L.) seedling roots with the high-throughput image analysis tool *ARIA* (Automatic Root Image Analysis)**. *PLoS ONE* 2014, 9(9).

45. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE: **A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species**. *PloS one* 2011, **6**(5):e19379.

46. Evanno G, Regnaut S, Goudet J: **Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study**. *Mol Ecol* 2005, **14**(8):2611-2620.

47. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES: **TASSEL: software for association mapping of complex traits in diverse samples**. *Bioinformatics* 2007, **23**(19):2633-2635.

48. Hardy OJ, Vekemans X: **SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels.** *Molecular Ecology Notes* 2002, 2: 618-620.

49. Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z: **GAPIT: genome association and prediction integrated tool**. *Bioinformatics* 2012, **28**(18):2397-2399.

50. Gao X, Starmer J, Martin ER: **A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms**. *Genet Epidemiol* 2008, **32**(4):361-369.

51. Johnson RC, Nelson GW, Troyer JL, Lautenberger JA, Kessing BD, Winkler CA, O'Brien SJ: **Accounting for multiple comparisons in a genome-wide association study (GWAS)**. *BMC Genomics* 2010, **11**:724.

Table 1. ***Trait designations and descriptions collected manually and by ARIA***

Trait Name	Symbol	Trait Description
Total Root Length	TRL	Cumulative length of all the roots in centimeters
Primary Root Length	PRL	Length of the Primary root in centimeters
Secondary Root Length	SEL	Cumulative length of all secondary roots in centimeters
Center of Point	COP	Absolute center of the root regardless of root length.
Maximum Number of Roots	MNR	The 84th percentile value of the sum of every row
Perimeter	PER	Total number of network pixels connected to a background pixel
Depth	DEP	The maximum vertical distance reached by the root system
Width	WID	The maximum horizontal width of the whole RSA
Width/Depth ratio	WDR	The ratio of the maximum width to depth
Median	MED	The median number of roots at all Y-location
Total Number of Roots	TNR	Total number of roots
Convex Area	CVA	The area of the convex hull that encloses the entire root image
Network Area	NWA	The number of pixels that are connected in the skeletonized image
Bushiness	BSH	The ratio of the maximum to the median number of roots.
Length Distribution	LED	The ratio of TRL in the upper one-third of the root to the TRL
Diameter	DIA	Diameter of the primary root
Surface Area	SUA	Surface area of the entire root system
Standard Root Length	SRL	Total root length divided by root volume
Shoot Length	SHL	Total Length of the shoot to the longest leaf tip in cm
Shoot Dry weight	SDW	Total dry weight of only the plant shoot
Root Dry Weight	RDW	Total dry weight of only the plant roots
Total Plant Biomass	TPB	Root dry weight and Shoot dry weight added together

Table 2. Trait statistics collected for all 22 traits. Line marks heritability cutoff for traits analyzed further

Trait	Mean	Std. Dev	Minimum	Maximum	H^2
TPB	0.107 g	0.036	0.016 g	0.253 g	0.491
WID	5.23	1.64	0.81	10.5	0.489
TNR	11.05	4.94	1	26.67	0.486
RDW	0.058 g	0.021	.005 g	0.145 g	0.479
SDW	0.049g	0.019	.005 g	0.124g	0.474
MED	5.12	2.61	1	16	0.449
COP	0.43	0.07	0.18	0.74	0.441
SHL	15.77 cm	4.42	2.55 cm	30.6 cm	0.431
SUA	10.22 cm^2	4.32	1.16 cm^2	25.04 cm^2	0.424
TRL	190.05 cm	98.07	16.39 cm	536.33 cm	0.423
SEL	149.32 cm	92.8	0.16 cm	490.59 cm	0.419
NWA	1.09	0.61	0.03	3.26	0.39
MNR	80.8	33.94	4	196	0.385
DIA	0.12	0.03	0.05	0.35	0.333
PER	143.38 cm	54.06	9.77 cm	307.07 cm	0.305
CVA	87.79	43.36	1.24	218.9	0.303
PRL	28.45 cm	8.35	4.09 cm	47.06 cm	0.281
WDR	0.25	0.42	0.08	13.01	0.268
DEP	24.17	6.54	3.56	34.88	0.257
SRL	0.59	0.4	0.05	2.54	0.209
LED	0.76	0.31	0.02	3.13	0.186
BSH	2.4	0.807	1	10	0.119

TPB = Total Plant Biomass, WID = Width, TNR = Total number of roots, RDW = Root Dry Weight, SDW= Shoot Dry Weight, MED = Median, COP = Center of Point, SHL= Shoot Length, SUA = Surface Area, TRL= Total Root Length, SEL= Secondary Root Length, NWA = Network Area, MNR= Maximum Root Number, DIA = Diameter, PER = Perimeter, CVA = Convex Root Area, PRL = Primary Root Length, WDR = Width Depth Ratio, DEP = Depth, SRL = Standard Root Length, LED = Length Distribution, BSH = Bushiness

Table 3. Pearson (r) correlations between all 22 traits collected

TRL	SUA	PRL	SEL	COP	MNR	PER	DEP	WID	WDR	MED	TNR	CVA	NWA	LED	DIA	SRL	BSH	SDW	RDW	TPB	SHL
1	0.898	0.716	0.983	0.117	0.391	0.776	0.718	0.632	-0.098	0.921	0.908	0.801	0.980	-0.262	0.346	-0.397	-0.224	0.647	0.635	0.708	0.701
	1	0.823	0.889	0.010	0.460	0.822	0.797	0.705	-0.166	0.804	0.781	0.856	0.900	-0.263	0.506	-0.526	-0.247	0.750	0.673	0.788	0.691
		1	0.682	-0.155	0.636	0.867	0.966	0.683	-0.241	0.548	0.507	0.868	0.728	-0.178	0.323	-0.410	-0.228	0.537	0.564	0.546	0.617
			1	0.135	0.352	0.756	0.682	0.632	-0.131	0.949	0.932	0.791	0.993	-0.254	0.332	-0.378	-0.259	0.635	0.618	0.695	0.682
				1	-0.005	-0.006	-0.195	-0.026	0.109	0.216	0.220	-0.035	0.097	-0.374	0.262	-0.213	-0.007	0.042	0.049	0.070	0.062
					1	0.711	0.614	0.357	-0.174	0.273	0.316	0.496	0.394	-0.236	0.351	-0.416	0.003	0.384	0.304	0.325	0.388
						1	0.837	0.731	-0.182	0.652	0.653	0.871	0.784	-0.240	0.408	-0.465	0.003	0.565	0.582	0.583	0.619
							1	0.609	-0.282	0.539	0.509	0.825	0.729	-0.127	0.230	-0.366	-0.159	0.520	0.551	0.522	0.608
								1	-0.004	0.572	0.568	0.866	0.641	-0.067	0.401	-0.360	-0.220	0.487	0.459	0.486	0.510
									1	-0.106	-0.071	-0.124	-0.143	0.016	0.000	-0.002	-0.104	-0.084	-0.169	-0.125	-0.145
										1	0.945	0.667	0.932	-0.291	0.319	-0.386	0.253	0.605	0.592	0.634	0.651
											1	0.638	0.921	-0.188	0.349	-0.396	-0.329	0.598	0.578	0.616	0.638
												1	0.822	-0.198	0.390	-0.410	-0.231	0.574	0.540	0.618	0.608
													1	-0.255	0.324	-0.379	-0.256	0.628	0.639	0.700	0.694
														1	-0.411	0.326	0.417	-0.178	-0.288	-0.249	-0.298
															1	-0.656	-0.003	0.400	0.364	0.415	0.304
																1	-0.011	-0.454	-0.387	-0.422	-0.360
																	1	-0.163	-0.173	-0.214	-0.186
																		1	0.548	0.851	0.769
																			1	0.859	0.409
																				1	0.654
																					1

TPB = Total Plant Biomass, WID = Width, TNR = Total number of roots, RDW = Root Dry Weight, SDW= Shoot Dry Weight, MED = Median, COP = Center of Point, SHL= Shoot Length, SUA = Surface Area, TRL= Total Root Length, SEL= Secondary Root Length, NWA = Network Area, MNR= Maximum Root Number, DIA = Diameter, PER = Perimeter, CVA = Convex Root Area, PRL = Primary Root Length, WDR = Width Depth Ratio, DEP = Depth, SRL = Standard Root Length, LED = Length Distribution, BSH = Bushiness

Table 4. Gene model absolute expression values found in B73 genome.

Gene Model	Absolute expression value in primary root at V1
GRMZM2G153722	7456.31
GRMZM2G053511	66.76
GRMZM2G002879	1216.56
GRMZM2G154864	4826.04
GRMZM2G070837	53.19
GRMZM2G095969	70.27
GRMZM2G322186	4784.54

Supplementary Table 1. ***List of all significant marker trait associations determined by GWAS using MLM and GLM***

GWAS Method	Trait	SNP	Chromosome	Position	P.value	Allelic effect
MLM	SRL	S3_49565840	3	49565840	3.21E-09	-0.15559836
MLM	SRL	S3_49619564	3	49619564	2.46E-08	0.140166308
MLM	SRL	S3_49619525	3	49619525	2.46E-08	-0.140166308
MLM	BSH	S2_234437741	2	234437741	1.75E-07	-0.208725017
GLM	MRN	S4_62787846	4	62787846	2.04E-07	NA
GLM	MED	S8_74681862	8	74681862	1.61E-07	NA
GLM	NWA	S2_21818271	2	21818271	4.46E-07	NA
GLM	NWA	S4_62787846	4	62787846	4.86E-07	NA
GLM	PER	S2_202635930	2	202635930	2.01E-09	NA
GLM	PER	S2_202176704	2	202176704	2.44E-08	NA
GLM	PER	S8_146152722	8	146152722	5.02E-08	NA
GLM	PER	S3_187226846	3	187226846	2.04E-07	NA
GLM	PER	S5_213079932	5	213079932	2.63E-07	NA
GLM	PER	S5_188609024	5	188609024	4.74E-07	NA
GLM	PER	S2_21818271	2	21818271	4.84E-07	NA
GLM	RDW	S2_201899235	2	201899235	2.67E-08	NA
GLM	RDW	S2_210962971	2	210962971	2.89E-08	NA
GLM	RDW	S9_116263697	9	116263697	5.31E-08	NA
GLM	RDW	S9_124316950	9	124316950	6.12E-08	NA
GLM	RDW	S9_139174871	9	139174871	1.18E-07	NA
GLM	RDW	S2_210978763	2	210978763	1.20E-07	NA
GLM	RDW	S2_202176704	2	202176704	1.33E-07	NA
GLM	RDW	S9_5455309	9	5455309	1.40E-07	NA
GLM	RDW	S2_201965660	2	201965660	1.61E-07	NA
GLM	RDW	S9_130090472	9	130090472	1.92E-07	NA
GLM	RDW	S3_20534762	3	20534762	2.21E-07	NA
GLM	RDW	S2_202641741	2	202641741	2.43E-07	NA
GLM	RDW	S5_199978118	5	199978118	2.90E-07	NA
GLM	RDW	S4_18676302	4	18676302	3.22E-07	NA
GLM	RDW	S8_155440143	8	155440143	3.87E-07	NA
GLM	RDW	S7_159921873	7	159921873	4.06E-07	NA
GLM	RDW	S7_159921877	7	159921877	4.06E-07	NA
GLM	RDW	S2_210978535	2	210978535	4.20E-07	NA
GLM	RDW	S6_146218985	6	146218985	4.64E-07	NA
GLM	RDW	S9_124031518	9	124031518	4.74E-07	NA
GLM	RDW	S2_41497406	2	41497406	4.79E-07	NA
GLM	RDW	S9_122595736	9	122595736	4.84E-07	NA
GLM	RDW	S9_122595739	9	122595739	4.84E-07	NA

Supplementary Table 1 continued						
GLM	RDW	S9_122595744	9	122595744	4.84E-07	NA
GLM	RDW	S2_212541278	2	212541278	5.08E-07	NA
GLM	SDW	S2_202178253	2	202178253	5.84E-10	NA
GLM	SDW	S8_79284127	8	79284127	2.55E-09	NA
GLM	SDW	S1_255084426	1	255084426	1.56E-08	NA
GLM	SDW	S3_167473718	3	167473718	7.54E-08	NA
GLM	SDW	S3_167362874	3	167362874	1.06E-07	NA
GLM	SDW	S1_198479489	1	198479489	1.13E-07	NA
GLM	SDW	S3_223090866	3	223090866	1.37E-07	NA
GLM	SDW	S6_129436495	6	129436495	1.59E-07	NA
GLM	SDW	S2_202635930	2	202635930	1.84E-07	NA
GLM	SDW	S2_204294242	2	204294242	3.40E-07	NA
GLM	SDW	S6_152308251	6	152308251	4.27E-07	NA
GLM	SDW	S9_103173310	9	103173310	4.57E-07	NA
GLM	SDW	S2_55894837	2	55894837	5.33E-07	NA
GLM	SHL	S2_55894837	2	55894837	2.51E-09	NA
GLM	SHL	S2_55878267	2	55878267	4.94E-09	NA
GLM	SHL	S2_55875101	2	55875101	9.86E-09	NA
GLM	SHL	S2_55878363	2	55878363	2.68E-08	NA
GLM	SHL	S2_55893351	2	55893351	3.24E-08	NA
GLM	SHL	S2_55893619	2	55893619	3.60E-08	NA
GLM	SHL	S2_55893782	2	55893782	9.42E-08	NA
GLM	SHL	S2_202178253	2	202178253	1.05E-07	NA
GLM	SHL	S1_12977206	1	12977206	1.33E-07	NA
GLM	SHL	S8_19120665	8	19120665	1.92E-07	NA
GLM	SHL	S8_18894046	8	18894046	1.97E-07	NA
GLM	SHL	S5_19842373	5	19842373	2.23E-07	NA
GLM	SHL	S9_128457179	9	128457179	3.36E-07	NA
GLM	SHL	S2_55875076	2	55875076	3.39E-07	NA
GLM	SHL	S5_174591302	5	174591302	4.22E-07	NA
GLM	TNR	S4_62787846	4	62787846	2.04E-07	NA
GLM	TRL	S4_62787846	4	62787846	3.67E-07	NA
GLM	PRL	S2_202635930	2	2.03E+08	3.73E-09	NA
GLM	PRL	S8_146152722	8	1.46E+08	3.77E-08	NA
GLM	PRL	S1_34401500	1	34401500	1.09E-07	NA
GLM	PRL	S6_7125119	6	7125119	1.17E-07	NA
GLM	PRL	S2_202176704	2	2.02E+08	1.33E-07	NA
GLM	PRL	S3_187226846	3	1.87E+08	2.05E-07	NA
GLM	PRL	S10_144957882	10	1.45E+08	2.42E-07	NA
GLM	PRL	S6_146218985	6	1.46E+08	3.19E-07	NA

Supplementary Table 1 continued						
GLM	PRL	S2_21818271	2	21818271	3.98E-07	NA
GLM	PRL	S2_202174949	2	2.02E+08	4.04E-07	NA
GLM	PRL	S2_21802146	2	21802146	5.03E-07	NA
GLM	SUA	S2_202635930	2	202635930	9.74E-10	NA
GLM	SUA	S3_190009541	3	190009541	6.12E-09	NA
GLM	SUA	S3_190009572	3	190009572	6.12E-09	NA
GLM	SUA	S2_21818271	2	21818271	8.42E-09	NA
GLM	SUA	S4_62787846	4	62787846	9.74E-09	NA
GLM	SUA	S2_202176704	2	202176704	1.31E-08	NA
GLM	SUA	S4_62568928	4	62568928	1.43E-08	NA
GLM	SUA	S6_146218985	6	146218985	1.97E-08	NA
GLM	SUA	S4_62573171	4	62573171	2.54E-08	NA
GLM	SUA	S4_62572919	4	62572919	2.66E-08	NA
GLM	SUA	S4_62412291	4	62412291	2.86E-08	NA
GLM	SUA	S4_62565527	4	62565527	3.18E-08	NA
GLM	SUA	S4_62565569	4	62565569	3.18E-08	NA
GLM	SUA	S2_202635915	2	202635915	3.37E-08	NA
GLM	SUA	S4_62695364	4	62695364	4.29E-08	NA
GLM	SUA	S2_202528876	2	202528876	4.87E-08	NA
GLM	SUA	S3_20534762	3	20534762	5.28E-08	NA
GLM	SUA	S4_62573079	4	62573079	5.28E-08	NA
GLM	SUA	S5_204385180	5	204385180	6.12E-08	NA
GLM	SUA	S4_62573001	4	62573001	7.17E-08	NA
GLM	SUA	S4_62353254	4	62353254	7.32E-08	NA
GLM	SUA	S4_62353279	4	62353279	7.32E-08	NA
GLM	SUA	S2_202178253	2	202178253	8.73E-08	NA
GLM	SUA	S2_202639340	2	202639340	9.41E-08	NA
GLM	SUA	S2_202528862	2	202528862	1.00E-07	NA
GLM	SUA	S3_187226846	3	187226846	1.01E-07	NA
GLM	SUA	S4_62694610	4	62694610	1.03E-07	NA
GLM	SUA	S2_212744456	2	212744456	1.07E-07	NA
GLM	SUA	S5_204522546	5	204522546	1.07E-07	NA
GLM	SUA	S4_62567488	4	62567488	1.17E-07	NA
GLM	SUA	S3_187450966	3	187450966	1.35E-07	NA
GLM	SUA	S2_212541278	2	212541278	1.42E-07	NA
GLM	SUA	S4_217095284	4	217095284	1.49E-07	NA
GLM	SUA	S2_13299683	2	13299683	1.50E-07	NA
GLM	SUA	S2_21802146	2	21802146	1.50E-07	NA
GLM	SUA	S4_62572909	4	62572909	1.60E-07	NA
GLM	SUA	S1_200409094	1	200409094	1.67E-07	NA

Supplementary Table 1 continued						
GLM	SUA	S4_62787629	4	62787629	1.68E-07	NA
GLM	SUA	S3_184267045	3	184267045	2.12E-07	NA
GLM	SUA	S4_62787622	4	62787622	2.22E-07	NA
GLM	SUA	S4_62573373	4	62573373	2.40E-07	NA
GLM	SUA	S4_62573339	4	62573339	2.59E-07	NA
GLM	SUA	S10_33221023	10	33221023	2.64E-07	NA
GLM	SUA	S4_62788968	4	62788968	2.77E-07	NA
GLM	SUA	S4_62564497	4	62564497	2.87E-07	NA
GLM	SUA	S4_56970265	4	56970265	3.15E-07	NA
GLM	SUA	S3_184267357	3	184267357	3.33E-07	NA
GLM	SUA	S4_63206226	4	63206226	3.40E-07	NA
GLM	SUA	S6_86049082	6	86049082	3.41E-07	NA
GLM	SUA	S4_62573340	4	62573340	3.51E-07	NA
GLM	SUA	S4_62573370	4	62573370	3.51E-07	NA
GLM	SUA	S1_200409145	1	200409145	3.99E-07	NA
GLM	SUA	S1_208781061	1	208781061	4.23E-07	NA
GLM	SUA	S4_142206837	4	142206837	4.44E-07	NA
GLM	SUA	S3_2154241	3	2154241	4.51E-07	NA
GLM	SUA	S3_190063251	3	190063251	4.67E-07	NA
GLM	SUA	S2_10375886	2	10375886	4.68E-07	NA
GLM	SUA	S2_27546469	2	27546469	4.90E-07	NA
GLM	TPB	S2_226586217	2	226586217	8.68E-09	NA
GLM	TPB	S2_202178253	2	202178253	1.43E-08	NA
GLM	TPB	S2_202528862	2	202528862	1.53E-08	NA
GLM	TPB	S1_255084426	1	255084426	2.99E-08	NA
GLM	TPB	S2_202635930	2	202635930	3.78E-08	NA
GLM	TPB	S1_179184476	1	179184476	5.03E-08	NA
GLM	TPB	S2_202176704	2	202176704	7.23E-08	NA
GLM	TPB	S6_129436495	6	129436495	9.51E-08	NA
GLM	TPB	S2_202641741	2	202641741	1.92E-07	NA
GLM	TPB	S9_139174871	9	139174871	2.11E-07	NA
GLM	TPB	S9_112489129	9	112489129	2.14E-07	NA
GLM	TPB	S2_204294242	2	204294242	2.37E-07	NA
GLM	TPB	S2_202739518	2	202739518	3.16E-07	NA
GLM	TPB	S9_130090472	9	130090472	3.52E-07	NA
GLM	TPB	S9_105339007	9	105339007	3.83E-07	NA
GLM	TPB	S9_105339008	9	105339008	3.83E-07	NA
GLM	TPB	S2_201965660	2	201965660	4.01E-07	NA
GLM	DIA	S2_7040758	2	7040758	5.34E-11	NA
GLM	DIA	S2_6850421	2	6850421	1.44E-10	NA

Supplementary Table 1 continued						
GLM	DIA	S2_6850420	2	6850420	2.19E-10	NA
GLM	DIA	S2_6172579	2	6172579	3.79E-10	NA
GLM	DIA	S3_2076703	3	2076703	4.83E-10	NA
GLM	DIA	S2_6994422	2	6994422	2.43E-09	NA
GLM	DIA	S2_13349810	2	13349810	2.63E-09	NA
GLM	DIA	S4_62568928	4	62568928	3.31E-09	NA
GLM	DIA	S4_62565527	4	62565527	3.55E-09	NA
GLM	DIA	S4_62565569	4	62565569	3.55E-09	NA
GLM	DIA	S2_6720877	2	6720877	3.62E-09	NA
GLM	DIA	S2_6720881	2	6720881	3.62E-09	NA
GLM	DIA	S2_6720882	2	6720882	3.62E-09	NA
GLM	DIA	S7_119838612	7	119838612	3.92E-09	NA
GLM	DIA	S7_119838613	7	119838613	3.92E-09	NA
GLM	DIA	S2_6333501	2	6333501	5.21E-09	NA
GLM	DIA	S4_62572919	4	62572919	5.21E-09	NA
GLM	DIA	S4_62573171	4	62573171	5.65E-09	NA
GLM	DIA	S2_3105910	2	3105910	6.06E-09	NA
GLM	DIA	S2_11361054	2	11361054	6.39E-09	NA
GLM	DIA	S2_4876527	2	4876527	6.80E-09	NA
GLM	DIA	S4_62787846	4	62787846	6.87E-09	NA
GLM	DIA	S2_12253075	2	12253075	7.56E-09	NA
GLM	DIA	S4_62573079	4	62573079	7.78E-09	NA
GLM	DIA	S2_10375886	2	10375886	1.19E-08	NA
GLM	DIA	S2_5836126	2	5836126	1.26E-08	NA
GLM	DIA	S2_5836127	2	5836127	1.26E-08	NA
GLM	DIA	S2_5836129	2	5836129	1.26E-08	NA
GLM	DIA	S2_5836131	2	5836131	1.26E-08	NA
GLM	DIA	S2_3106026	2	3106026	1.31E-08	NA
GLM	DIA	S2_3106027	2	3106027	1.31E-08	NA
GLM	DIA	S2_2811155	2	2811155	1.44E-08	NA
GLM	DIA	S4_62353254	4	62353254	1.48E-08	NA
GLM	DIA	S4_62353279	4	62353279	1.48E-08	NA
GLM	DIA	S4_62567488	4	62567488	1.50E-08	NA
GLM	DIA	S2_6460559	2	6460559	1.55E-08	NA
GLM	DIA	S2_2285818	2	2285818	1.67E-08	NA
GLM	DIA	S2_1267098	2	1267098	2.02E-08	NA
GLM	DIA	S4_62412291	4	62412291	2.14E-08	NA
GLM	DIA	S2_6471845	2	6471845	2.15E-08	NA
GLM	DIA	S4_62572909	4	62572909	2.62E-08	NA
GLM	DIA	S2_1654161	2	1654161	2.75E-08	NA

Supplementary Table 1 continued						
GLM	DIA	S2_1267306	2	1267306	2.93E-08	NA
GLM	DIA	S2_6177436	2	6177436	3.29E-08	NA
GLM	DIA	S2_2540702	2	2540702	3.61E-08	NA
GLM	DIA	S2_6466394	2	6466394	3.77E-08	NA
GLM	DIA	S2_3181711	2	3181711	4.01E-08	NA
GLM	DIA	S2_3181719	2	3181719	4.01E-08	NA
GLM	DIA	S4_62694610	4	62694610	4.16E-08	NA
GLM	DIA	S2_2593357	2	2593357	4.46E-08	NA
GLM	DIA	S4_62573339	4	62573339	4.49E-08	NA
GLM	DIA	S2_4990216	2	4990216	4.55E-08	NA
GLM	DIA	S2_6983715	2	6983715	5.77E-08	NA
GLM	DIA	S4_62412381	4	62412381	6.02E-08	NA
GLM	DIA	S2_4343433	2	4343433	6.76E-08	NA
GLM	DIA	S2_9315360	2	9315360	6.92E-08	NA
GLM	DIA	S4_62412374	4	62412374	7.10E-08	NA
GLM	DIA	S2_7040865	2	7040865	7.15E-08	NA
GLM	DIA	S2_7040867	2	7040867	7.15E-08	NA
GLM	DIA	S2_3106182	2	3106182	7.56E-08	NA
GLM	DIA	S2_6458690	2	6458690	7.60E-08	NA
GLM	DIA	S2_6471718	2	6471718	7.91E-08	NA
GLM	DIA	S2_6720880	2	6720880	8.10E-08	NA
GLM	DIA	S4_62564497	4	62564497	8.24E-08	NA
GLM	DIA	S2_6773738	2	6773738	8.52E-08	NA
GLM	DIA	S2_6355263	2	6355263	8.73E-08	NA
GLM	DIA	S2_7733390	2	7733390	9.20E-08	NA
GLM	DIA	S2_212525890	2	212525890	1.01E-07	NA
GLM	DIA	S3_205235096	3	205235096	1.01E-07	NA
GLM	DIA	S5_204385180	5	204385180	1.06E-07	NA
GLM	DIA	S2_3106513	2	3106513	1.11E-07	NA
GLM	DIA	S2_3106516	2	3106516	1.11E-07	NA
GLM	DIA	S4_62573340	4	62573340	1.11E-07	NA
GLM	DIA	S4_62573370	4	62573370	1.11E-07	NA
GLM	DIA	S6_144123661	6	144123661	1.15E-07	NA
GLM	DIA	S2_7190543	2	7190543	1.18E-07	NA
GLM	DIA	S10_37942464	10	37942464	1.19E-07	NA
GLM	DIA	S3_205392941	3	205392941	1.24E-07	NA
GLM	DIA	S2_1463303	2	1463303	1.25E-07	NA
GLM	DIA	S3_2118557	3	2118557	1.25E-07	NA
GLM	DIA	S1_32601179	1	32601179	1.28E-07	NA
GLM	DIA	S8_16444572	8	16444572	1.32E-07	NA

Supplementary Table 1 continued						
GLM	DIA	S2_202635930	2	202635930	1.51E-07	NA
GLM	DIA	S8_16444445	8	16444445	1.56E-07	NA
GLM	DIA	S8_16444587	8	16444587	1.57E-07	NA
GLM	DIA	S5_204522546	5	204522546	1.58E-07	NA
GLM	DIA	S3_136165588	3	136165588	1.62E-07	NA
GLM	DIA	S4_62412329	4	62412329	1.63E-07	NA
GLM	DIA	S10_4693744	10	4693744	1.70E-07	NA
GLM	DIA	S6_160037464	6	160037464	1.78E-07	NA
GLM	DIA	S3_219856818	3	219856818	1.79E-07	NA
GLM	DIA	S3_2071609	3	2071609	2.08E-07	NA
GLM	DIA	S4_62573001	4	62573001	2.09E-07	NA
GLM	DIA	S2_49707173	2	49707173	2.11E-07	NA
GLM	DIA	S2_7733127	2	7733127	2.16E-07	NA
GLM	DIA	S2_7733148	2	7733148	2.16E-07	NA
GLM	DIA	S2_3097857	2	3097857	2.29E-07	NA
GLM	DIA	S7_122076997	7	122076997	2.44E-07	NA
GLM	DIA	S1_100825248	1	100825248	2.56E-07	NA
GLM	DIA	S2_2628047	2	2628047	2.63E-07	NA
GLM	DIA	S2_202639340	2	202639340	2.72E-07	NA
GLM	DIA	S2_6994128	2	6994128	2.76E-07	NA
GLM	DIA	S7_122077001	7	122077001	2.77E-07	NA
GLM	DIA	S3_202051330	3	202051330	2.88E-07	NA
GLM	DIA	S2_4876525	2	4876525	2.88E-07	NA
GLM	DIA	S2_9011328	2	9011328	2.92E-07	NA
GLM	DIA	S2_13349544	2	13349544	2.95E-07	NA
GLM	DIA	S7_11476601	7	11476601	2.99E-07	NA
GLM	DIA	S4_62412368	4	62412368	3.15E-07	NA
GLM	DIA	S3_20534762	3	20534762	3.16E-07	NA
GLM	DIA	S2_3866630	2	3866630	3.21E-07	NA
GLM	DIA	S2_3866645	2	3866645	3.21E-07	NA
GLM	DIA	S3_184237812	3	184237812	3.22E-07	NA
GLM	DIA	S3_2067782	3	2067782	3.37E-07	NA
GLM	DIA	S2_202635915	2	202635915	3.42E-07	NA
GLM	DIA	S2_3773762	2	3773762	3.42E-07	NA
GLM	DIA	S4_62695364	4	62695364	3.47E-07	NA
GLM	DIA	S2_6177736	2	6177736	3.48E-07	NA
GLM	DIA	S2_7743653	2	7743653	3.50E-07	NA
GLM	DIA	S8_17059625	8	17059625	3.79E-07	NA
GLM	DIA	S2_13299683	2	13299683	3.93E-07	NA
GLM	DIA	S2_9119197	2	9119197	4.03E-07	NA

Supplementary Table 1 continued						
GLM	DIA	S2_3104647	2	3104647	4.09E-07	NA
GLM	DIA	S1_193164284	1	193164284	4.11E-07	NA
GLM	DIA	S2_202528876	2	202528876	4.27E-07	NA
GLM	DIA	S3_2767231	3	2767231	4.34E-07	NA
GLM	DIA	S4_121563375	4	121563375	4.38E-07	NA
GLM	DIA	S8_37808019	8	37808019	4.46E-07	NA
GLM	DIA	S2_12011959	2	12011959	4.68E-07	NA
GLM	DIA	S1_281950731	1	281950731	4.68E-07	NA
GLM	DIA	S1_253308208	1	253308208	4.79E-07	NA
GLM	DIA	S6_139619665	6	139619665	5.04E-07	NA
GLM	DIA	S2_3795762	2	3795762	5.14E-07	NA
GLM	DIA	S2_4891051	2	4891051	5.19E-07	NA
GLM	DIA	S8_17059371	8	17059371	5.23E-07	NA

Supplementary Table 2. ***All lines included in 384 Ames Panel Association mapping population.***					
IVP	**IVNO**	**Genotype name**	**TAXON**	**COUNTRY**	**STATE**
Ames	2332	Bei 10 = North 10	Zea mays subsp. mays	China	
Ames	2336	52220	Zea mays subsp. mays	China	
Ames	2523	38-11R PARENT HB 19 INB	Zea mays subsp. mays	Portugal	
Ames	14115	Va35C	Zea mays subsp. mays	United States	Virginia
Ames	14116	Va36A	Zea mays subsp. mays	United States	Virginia
Ames	19000	VaW6	Zea mays subsp. mays	United States	Virginia
Ames	19008	Va24	Zea mays subsp. mays	United States	Virginia
Ames	19010	Va37	Zea mays subsp. mays	United States	Virginia
Ames	19011	Va38	Zea mays subsp. mays	United States	Virginia
Ames	19012	Va39	Zea mays subsp. mays	United States	Virginia
Ames	19013	Va46	Zea mays subsp. mays	United States	Virginia
Ames	19016	Va59	Zea mays subsp. mays	United States	Virginia
Ames	19019	Va91	Zea mays subsp. mays	United States	Virginia
Ames	19293	Wf9	Zea mays subsp. mays	United States	Indiana
Ames	19308	A634	Zea mays subsp. mays	United States	Minnesota
Ames	19313	C123	Zea mays subsp. mays	United States	Connecticut
Ames	19318	H107	Zea mays subsp. mays	United States	Indiana
Ames	19319	H95	Zea mays subsp. mays	United States	Indiana
Ames	19326	R168	Zea mays subsp. mays	United States	Illinois
Ames	19327	Tx303	Zea mays subsp. mays	United States	Texas
Ames	19328	Va22	Zea mays subsp. mays	United States	Virginia
Ames	20119	Mo40	Zea mays subsp. mays	United States	Missouri
Ames	20137	H25W	Zea mays subsp. mays	United States	Indiana
Ames	22016	C15	Zea mays subsp. mays	United States	Connecticut
Ames	22017	C18	Zea mays subsp. mays	United States	Connecticut
Ames	23410	A265	Zea mays subsp. mays	United States	Minnesota
Ames	23413	A286	Zea mays subsp. mays	United States	Minnesota
Ames	23435	A427	Zea mays subsp. mays	United States	Minnesota
Ames	23456	A617	Zea mays subsp. mays	United States	Minnesota
Ames	23466	A630	Zea mays subsp. mays	United States	Minnesota
Ames	23471	A633	Zea mays subsp. mays	United States	Minnesota
Ames	23474	A636	Zea mays subsp. mays	United States	Minnesota
Ames	23475	A637	Zea mays subsp. mays	United States	Minnesota
Ames	23478	A643	Zea mays subsp. mays	United States	Minnesota
Ames	23479	A644	Zea mays subsp. mays	United States	Minnesota
Ames	23480	A645	Zea mays subsp. mays	United States	Minnesota
Ames	24705	MS4	Zea mays subsp. mays	United States	Michigan
Ames	24711	MS68	Zea mays subsp. mays	United States	Michigan
Ames	24713	MS72	Zea mays subsp. mays	United States	Michigan

Supplementary Table 2 continued					
Ames	24716	MS76	Zea mays subsp. mays	United States	Michigan
Ames	24718	MS78	Zea mays subsp. mays	United States	Michigan
Ames	24720	MS80	Zea mays subsp. mays	United States	Michigan
Ames	24723	MS91	Zea mays subsp. mays	United States	Michigan
Ames	24727	MS106	Zea mays subsp. mays	United States	Michigan
Ames	24730	MS132	Zea mays subsp. mays	United States	Michigan
Ames	24732	MS141	Zea mays subsp. mays	United States	Michigan
Ames	24735	MS198	Zea mays subsp. mays	United States	Michigan
Ames	24747	MS222	Zea mays subsp. mays	United States	Michigan
Ames	24748	MS223	Zea mays subsp. mays	United States	Michigan
Ames	24749	MS224	Zea mays subsp. mays	United States	Michigan
Ames	24751	MS226	Zea mays subsp. mays	United States	Michigan
Ames	24989	Va99	Zea mays subsp. mays	United States	Virginia
Ames	25372	Pa91HT1	Zea mays subsp. mays	United States	Pennsylvania
Ames	26021	P8	Zea mays subsp. mays	United States	Indiana
Ames	26120	CI 20	Zea mays subsp. mays	United States	Missouri
Ames	26743	WX38-11	Zea mays subsp. mays	United States	Iowa
Ames	26774	H14	Zea mays subsp. mays	United States	Indiana
Ames	26775	H19	Zea mays subsp. mays	United States	Indiana
Ames	26776	H22w	Zea mays subsp. mays	United States	Indiana
Ames	26777	H23w	Zea mays subsp. mays	United States	Indiana
Ames	26778	H26w	Zea mays subsp. mays	United States	Indiana
Ames	26779	H27w	Zea mays subsp. mays	United States	Indiana
Ames	26781	H29w	Zea mays subsp. mays	United States	Indiana
Ames	26783	H41	Zea mays subsp. mays	United States	Indiana
Ames	26788	H50	Zea mays subsp. mays	United States	Indiana
Ames	26790	H52	Zea mays subsp. mays	United States	Indiana
Ames	26791	H55	Zea mays subsp. mays	United States	Indiana
Ames	26792	H59	Zea mays subsp. mays	United States	Indiana
Ames	26795	H88	Zea mays subsp. mays	United States	Indiana
Ames	26909	Mo41	Zea mays subsp. mays	United States	Missouri
Ames	27017	CH705-8	Zea mays subsp. mays	Canada	Ontario
Ames	27018	CH711-10	Zea mays subsp. mays	Canada	Ontario
Ames	27019	CH732-12	Zea mays subsp. mays	Canada	Ontario
Ames	27020	CH741-6	Zea mays subsp. mays	Canada	Ontario
Ames	27069	CH701-30	Zea mays subsp. mays	Canada	Ontario
Ames	27122	K148	Zea mays subsp. mays	United States	Kansas
Ames	27124	Ki11	Zea mays subsp. mays	Thailand	
Ames	27125	Ki21	Zea mays subsp. mays	Thailand	
Ames	27136	Mo.G	Zea mays subsp. mays	United States	N. Carolina

Supplementary Table 2 continued					
Ames	27140	NC260	Zea mays subsp. mays	United States	N. Carolina
Ames	27149	NC306	Zea mays subsp. mays	United States	N. Carolina
Ames	27150	NC308	Zea mays subsp. mays	United States	N. Carolina
Ames	27188	SA24	Zea mays subsp. mays	United States	N. Carolina
Ames	27193	Va85	Zea mays subsp. mays	United States	Virginia
Ames	27444	Il731a	Zea mays subsp. mays	United States	Illinois
Ames	27445	Il767b	Zea mays subsp. mays	United States	Illinois
Ames	28186	P39 Goodman-Buckler	Zea mays subsp. mays	United States	Indiana
Ames	28360	Mo401	Zea mays subsp. mays	United States	Missouri
Ames	28361	Mo402	Zea mays subsp. mays	United States	Missouri
Ames	28366	N7A Goodman-Buckler	Zea mays subsp. mays	United States	Nebraska
Ames	28930	Mo30W	Zea mays subsp. mays	United States	Missouri
Ames	28935	Mo37	Zea mays subsp. mays	United States	Missouri
Ames	28937	Mo39	Zea mays subsp. mays	United States	Missouri
NSL	22630	K150	Zea mays subsp. mays	United States	Kansas
NSL	22635	K41	Zea mays subsp. mays	United States	Kansas
NSL	28966	Oh40B	Zea mays subsp. mays	United States	Ohio
NSL	28968	OH84	Zea mays subsp. mays	United States	Ohio
NSL	29317	R221	Zea mays subsp. mays	United States	Illinois
NSL	30053	W22	Zea mays subsp. mays	United States	Wisconsin
NSL	30060	W23	Zea mays subsp. mays	United States	Wisconsin
NSL	30064	W24	Zea mays subsp. mays	United States	Wisconsin
NSL	30071	W32	Zea mays subsp. mays	United States	Wisconsin
NSL	30835	SD10	Zea mays subsp. mays	United States	South Dakota
NSL	30863	L	Zea mays subsp. mays	United States	Illinois
NSL	30868	R30	Zea mays subsp. mays	United States	Illinois
NSL	30880	R105	Zea mays subsp. mays	United States	Illinois
NSL	30903	90	Zea mays subsp. mays	United States	Illinois
NSL	30905	5120B	Zea mays subsp. mays	United States	Illinois
NSL	32734	ND408	Zea mays subsp. mays	United States	North Dakota
NSL	32736	ND480	Zea mays subsp. mays	United States	North Dakota
NSL	65865	B10	Zea mays subsp. mays	United States	Iowa
NSL	67792	Mo307ae	Zea mays subsp. mays	United States	Missouri
NSL	75976	IA DS 61	Zea mays subsp. mays	United States	Iowa
NSL	81598	A657	Zea mays subsp. mays	United States	Minnesota
NSL	197104	H116	Zea mays subsp. mays	United States	Indiana
NSL	437893	AusTRCF 305819	Zea mays subsp. mays	Australia	Queensland
NSL	437896	AusTRCF 305822	Zea mays subsp. mays	Australia	Queensland
NSL	437907	AusTRCF 305833	Zea mays subsp. mays	Australia	Queensland
NSL	437909	AusTRCF 305835	Zea mays subsp. mays	Australia	Queensland

Supplementary Table 2 continued					
NSL	437910	AusTRCF 305836	Zea mays subsp. mays	Australia	Queensland
NSL	437913	AusTRCF 305839	Zea mays subsp. mays	Australia	Queensland
NSL	437923	AusTRCF 305849	Zea mays subsp. mays	Australia	Queensland
NSL	437925	AusTRCF 306065	Zea mays subsp. mays	Australia	Queensland
NSL	437930	AusTRCF 306235	Zea mays subsp. mays	Australia	Queensland
NSL	437931	AusTRCF 306236	Zea mays subsp. mays	Australia	New S. Wales
NSL	437932	AusTRCF 306237	Zea mays subsp. mays	Australia	New S. Wales
NSL	437934	AusTRCF 306239	Zea mays subsp. mays	Australia	New S. Wales
NSL	437935	AusTRCF 306240	Zea mays subsp. mays	Australia	New S. Wales
NSL	437936	AusTRCF 306241	Zea mays subsp. mays	Australia	New S. Wales
NSL	437939	AusTRCF 306244	Zea mays subsp. mays	Australia	Queensland
NSL	437943	AusTRCF 306254	Zea mays subsp. mays	Australia	Queensland
NSL	437946	AusTRCF 306257	Zea mays subsp. mays	Australia	Queensland
NSL	437950	AusTRCF 306261	Zea mays subsp. mays	Australia	Queensland
NSL	437952	AusTRCF 306264	Zea mays subsp. mays	Australia	Queensland
NSL	437959	AusTRCF 306273	Zea mays subsp. mays	Australia	Queensland
NSL	437960	AusTRCF 306274	Zea mays subsp. mays	Australia	New S. Wales
NSL	437962	AusTRCF 306276	Zea mays subsp. mays	Australia	Queensland
NSL	437964	AusTRCF 306278	Zea mays subsp. mays	Australia	Queensland
NSL	437966	AusTRCF 306280	Zea mays subsp. mays	Australia	Queensland
NSL	437967	AusTRCF 306281	Zea mays subsp. mays	Australia	Queensland
NSL	437968	AusTRCF 306282	Zea mays subsp. mays	Australia	Queensland
NSL	437971	AusTRCF 306285	Zea mays subsp. mays	Australia	Queensland
NSL	437973	AusTRCF 306287	Zea mays subsp. mays	Australia	Queensland
NSL	437976	AusTRCF 306290	Zea mays subsp. mays	Australia	Queensland
NSL	437979	AusTRCF 306293	Zea mays subsp. mays	Australia	Queensland
NSL	437982	AusTRCF 306296	Zea mays subsp. mays	Australia	Queensland
NSL	437989	AusTRCF 306303	Zea mays subsp. mays	Australia	Queensland
NSL	437990	AusTRCF 306304	Zea mays subsp. mays	Australia	Queensland
NSL	437992	AusTRCF 306306	Zea mays subsp. mays	Australia	Queensland
NSL	437993	AusTRCF 306307	Zea mays subsp. mays	Australia	Queensland
NSL	437994	AusTRCF 306308	Zea mays subsp. mays	Australia	Queensland
NSL	437995	AusTRCF 306309	Zea mays subsp. mays	Australia	Queensland
NSL	437996	AusTRCF 306310	Zea mays subsp. mays	Australia	Queensland
NSL	438007	AusTRCF 306321	Zea mays subsp. mays	Australia	Queensland
NSL	438009	AusTRCF 306323	Zea mays subsp. mays	Australia	Queensland

Supplementary Table 2 continued					
NSL	438010	AusTRCF 306324	Zea mays subsp. mays	Australia	Queensland
NSL	438019	AusTRCF 306333	Zea mays subsp. mays	Australia	Queensland
NSL	438021	AusTRCF 306335	Zea mays subsp. mays	Australia	Queensland
NSL	438022	AusTRCF 306336	Zea mays subsp. mays	Australia	Queensland
NSL	438023	AusTRCF 306337	Zea mays subsp. mays	Australia	Queensland
NSL	438029	AusTRCF 306343	Zea mays subsp. mays	Australia	Queensland
NSL	438030	AusTRCF 306344	Zea mays subsp. mays	Australia	Queensland
NSL	438031	AusTRCF 306345	Zea mays subsp. mays	Australia	Queensland
NSL	438033	AusTRCF 306347	Zea mays subsp. mays	Australia	Queensland
NSL	438034	AusTRCF 306348	Zea mays subsp. mays	Australia	Queensland
NSL	438036	AusTRCF 306350	Zea mays subsp. mays	Australia	Queensland
NSL	438038	AusTRCF 306352	Zea mays subsp. mays	Australia	Queensland
PI	186182	INBRED 378	Zea mays subsp. mays	Uruguay	
PI	186185	INBRED 605	Zea mays subsp. mays	Uruguay	
PI	186190	INBRED 624	Zea mays subsp. mays	Uruguay	
PI	186192	INBRED 45	Zea mays subsp. mays	Australia	
PI	186193	INBRED A-243-1	Zea mays subsp. mays	South Africa	
PI	186199	INBRED 141	Zea mays subsp. mays	Australia	
PI	186215	INBRED 2-687	Zea mays subsp. mays	Argentina	
PI	186216	INBRED 1-1265	Zea mays subsp. mays	Argentina	
PI	186217	INBRED 19-86	Zea mays subsp. mays	Argentina	
PI	186218	INBRED 34-1141	Zea mays subsp. mays	Argentina	
PI	186220	INBRED 34-1196	Zea mays subsp. mays	Argentina	
PI	186226	INBRED 305	Zea mays subsp. mays	Uruguay	
PI	186227	INBRED 309	Zea mays subsp. mays	Uruguay	
PI	186229	INBRED 321	Zea mays subsp. mays	Uruguay	
PI	186230	INBRED 334	Zea mays subsp. mays	Uruguay	
PI	198888	4F-35 BK	Zea mays subsp. mays	Argentina	
PI	198890	4F-203 AM 6	Zea mays subsp. mays	Argentina	
PI	198892	4F-234 BX 4	Zea mays subsp. mays	Argentina	
PI	198895	4F-285 TX 15	Zea mays subsp. mays	Argentina	
PI	198897	4F-306 108	Zea mays subsp. mays	Argentina	
PI	198902	4F-345 CN 12	Zea mays subsp. mays	Argentina	
PI	200179	NY 3 (Neveh Yaar)	Zea mays subsp. mays	Israel	
PI	200182	NY 159 (Neveh Yaar)	Zea mays subsp. mays	Israel	
PI	200184	NY 166 (Neveh Yaar)	Zea mays subsp. mays	Israel	
PI	200185	NY 188 (Neveh Yaar)	Zea mays subsp. mays	Israel	
PI	200187	NY 318 (Nevey Yaar)	Zea mays subsp. mays	Israel	
PI	200188	NY 364 (Neveh Yaar)	Zea mays subsp. mays	Israel	
PI	200193	NY 643 (Neveh Yaar)	Zea mays subsp. mays	Israel	

Supplementary Table 2 continued					
PI	200194	NY 971 (Neveh Yaar)	Zea mays subsp. mays	Israel	
PI	200196	NY 1000 (Neveh Yaar)	Zea mays subsp. mays	Israel	
PI	221734	A14 INBRED (POTCHEFSTROOM PEARL)	Zea mays subsp. mays	South Africa	Transvaal
PI	221735	A15-1 INBRED (POTCHEFSTROOM PEARL)	Zea mays subsp. mays	South Africa	Transvaal
PI	221736	A16-3-2 INBRED (POTCHEFSTROOM PEARL)	Zea mays subsp. mays	South Africa	Transvaal
PI	221747	E205-1-1-1 INBRED (S5 SYN. ANVELD)	Zea mays subsp. mays	South Africa	Transvaal
PI	221773	A415-1-3 INBRED	Zea mays subsp. mays	South Africa	Transvaal
PI	221775	A436-1 INBRED	Zea mays subsp. mays	South Africa	Transvaal
PI	221789	E683-1-2-1(S5) INBRED	Zea mays subsp. mays	South Africa	Transvaal
PI	221790	E684-1-1-1(S5) INBRED	Zea mays subsp. mays	South Africa	Transvaal
PI	221804	A242-2(S10) INBRED (PERUVIAN)	Zea mays subsp. mays	South Africa	Transvaal
PI	221805	A243-1-2(S10) INBRED (PERUVIAN)	Zea mays subsp. mays	South Africa	Transvaal
PI	221806	A256-1(S10) INBRED (PERUVIAN)	Zea mays subsp. mays	South Africa	Transvaal
PI	221811	A302-1-2(S10) INBRED (SERVENTINA)	Zea mays subsp. mays	South Africa	Transvaal
PI	221813	A325-1(S10) INBRED (HOTNOT)	Zea mays subsp. mays	South Africa	Transvaal
PI	221820	C410-1(F11) INBRED (HOTNOT CROSSES)	Zea mays subsp. mays	South Africa	Transvaal
PI	257514	FV181	Zea mays subsp. mays	France	
PI	257517	FC46	Zea mays subsp. mays	France	
PI	267171	T8445 INBRED	Zea mays subsp. mays	Former Soviet Union	
PI	303925	NO. 1004 INBRED	Zea mays subsp. mays	Spain	
PI	303926	NO. 1019 INBRED	Zea mays subsp. mays	Spain	
PI	303928	NO. 1032 INBRED	Zea mays subsp. mays	Spain	
PI	303929	NO. 1037 INBRED	Zea mays subsp. mays	Spain	
PI	303930	NO. 1049 INBRED	Zea mays subsp. mays	Spain	
PI	303932	NO. 1068 INBRED	Zea mays subsp. mays	Spain	
PI	303933	NO. 1070 INBRED	Zea mays subsp. mays	Spain	
PI	303936	NO. 1174 INBRED	Zea mays subsp. mays	Spain	
PI	303940	NO. 1201 INBRED	Zea mays subsp. mays	Spain	
PI	303943	TN 53-1-2	Zea mays subsp. mays	Taiwan	
PI	340812	NY 121 (Neveh Yaar)	Zea mays subsp. mays	Israel	

Supplementary Table 2 continued					
PI	340813	NY 123 (Neveh Yaar)	Zea mays subsp. mays	Israel	
PI	340817	G3 T5	Zea mays subsp. mays	Romania	
PI	340821	G22 T122	Zea mays subsp. mays	Romania	
PI	340823	G14 T133	Zea mays subsp. mays	Romania	
PI	340824	G15 T134	Zea mays subsp. mays	Romania	
PI	340827	T141	Zea mays subsp. mays	Romania	
PI	340875	IA DS 43-W	Zea mays subsp. mays	United States	Iowa
PI	391660	CHI-41	Zea mays subsp. mays	China	Shaanxi
PI	405705	CHAN 11 INBRED	Zea mays subsp. mays	China	
PI	405711	BAI TOU SHUANG IN.(JI 095	Zea mays subsp. mays	China	
PI	406106	A14NW	Zea mays subsp. mays	South Africa	
PI	406107	A57N	Zea mays subsp. mays	South Africa	
PI	406108	A98NW	Zea mays subsp. mays	South Africa	
PI	406110	A178N	Zea mays subsp. mays	South Africa	
PI	406123	A579N	Zea mays subsp. mays	South Africa	
PI	406124	A622N	Zea mays subsp. mays	South Africa	
PI	406125	A641N	Zea mays subsp. mays	South Africa	
PI	406127	A664N	Zea mays subsp. mays	South Africa	
PI	415088	4581 INBRED	Zea mays subsp. mays	Hungary	
PI	506411	M6411	Zea mays subsp. mays	United States	Oklahoma
PI	506412	M6415	Zea mays subsp. mays	United States	Oklahoma
PI	506413	M6421	Zea mays subsp. mays	United States	Oklahoma
PI	508277	SD42	Zea mays subsp. mays	United States	South Dakota
PI	511309	NC252	Zea mays subsp. mays	United States	North Carolina
PI	511310	NC254	Zea mays subsp. mays	United States	North Carolina
PI	511311	NC256	Zea mays subsp. mays	United States	North Carolina
PI	517973	Pa879	Zea mays subsp. mays	United States	Pennsylvania
PI	517974	Pa880	Zea mays subsp. mays	United States	Pennsylvania
PI	524970	SD46	Zea mays subsp. mays	United States	South Dakota
PI	531081	Pa356	Zea mays subsp. mays	United States	Pennsylvania
PI	531082	Pa376	Zea mays subsp. mays	United States	Pennsylvania
PI	531085	NC262	Zea mays subsp. mays	United States	N. Carolina
PI	537097	LH195	Zea mays subsp. mays	United States	
PI	537099	LH205	Zea mays subsp. mays	United States	
PI	538010	LH206	Zea mays subsp. mays	United States	
PI	538011	LH220Ht	Zea mays subsp. mays	United States	
PI	538229	SD53	Zea mays subsp. mays	United States	South Dakota

Supplementary Table 2 continued					
PI	538242	SD106	Zea mays subsp. mays	United States	South Dakota
PI	538244	SD108	Zea mays subsp. mays	United States	South Dakota
PI	539924	LH202	Zea mays subsp. mays	United States	
PI	542716	NP87	Zea mays subsp. mays	United States	Nebraska
PI	542777	HP72-11	Zea mays subsp. mays	United States	Indiana
PI	542955	Va4	Zea mays subsp. mays	United States	Virginia
PI	542956	Va5	Zea mays subsp. mays	United States	Virginia
PI	547088	LH208	Zea mays subsp. mays	United States	
PI	550442	Mo20W	Zea mays subsp. mays	United States	Missouri
PI	550469	B46	Zea mays subsp. mays	United States	Iowa
PI	550473	B73	Zea mays subsp. mays	United States	Iowa
PI	550496	H102	Zea mays subsp. mays	United States	Indiana
PI	550497	H103	Zea mays subsp. mays	United States	Indiana
PI	550527	H111	Zea mays subsp. mays	United States	Indiana
PI	550555	NC250	Zea mays subsp. mays	United States	N. Carolina
PI	550558	DE811	Zea mays subsp. mays	United States	Delaware
PI	550903	89199	Zea mays subsp. mays	Cameroon	
PI	558520	Mo1W	Zea mays subsp. mays	United States	Missouri
PI	558521	Mo2RF	Zea mays subsp. mays	United States	Missouri
PI	558532	Mo17	Zea mays subsp. mays	United States	Missouri
PI	559380	ICI 193	Zea mays subsp. mays	United States	
PI	559381	ICI 441	Zea mays subsp. mays	United States	
PI	559382	ICI 740	Zea mays subsp. mays	United States	
PI	559383	ICI 893	Zea mays subsp. mays	United States	
PI	559918	NQ508	Zea mays subsp. mays	United States	Illinois
PI	561694	NYRD4058	Zea mays subsp. mays	United States	New York
PI	568158	N199	Zea mays subsp. mays	United States	Nebraska
PI	572413	Oh599	Zea mays subsp. mays	United States	Ohio
PI	583352	Mo47	Zea mays subsp. mays	United States	Missouri
PI	583846	H126W	Zea mays subsp. mays	United States	Indiana
PI	587126	C13	Zea mays subsp. mays	United States	Connecticut
PI	587127	H105W	Zea mays subsp. mays	United States	Indiana
PI	587128	H84	Zea mays subsp. mays	United States	Indiana
PI	587131	HP301	Zea mays subsp. mays	United States	Indiana
PI	587138	A554	Zea mays subsp. mays	United States	Minnesota
PI	587140	A632	Zea mays subsp. mays	United States	Minnesota
PI	587150	Va35	Zea mays subsp. mays	United States	Virginia
PI	592735	R230	Zea mays subsp. mays	United States	Illinois
PI	593009	Hi27	Zea mays subsp. mays	United States	Hawaii
PI	593015	Hi34	Zea mays subsp. mays	United States	Hawaii

Supplementary Table 2 continued					
PI	594050	N501	Zea mays subsp. mays	United States	Nebraska
PI	594051	N502	Zea mays subsp. mays	United States	Nebraska
PI	594058	N509	Zea mays subsp. mays	United States	Nebraska
PI	594059	N510	Zea mays subsp. mays	United States	Nebraska
PI	594060	N511	Zea mays subsp. mays	United States	Nebraska
PI	594061	N512	Zea mays subsp. mays	United States	Nebraska
PI	594063	N514	Zea mays subsp. mays	United States	Nebraska
PI	594064	N515	Zea mays subsp. mays	United States	Nebraska
PI	594065	N516	Zea mays subsp. mays	United States	Nebraska
PI	594066	N517	Zea mays subsp. mays	United States	Nebraska
PI	594067	N518	Zea mays subsp. mays	United States	Nebraska
PI	594070	N521	Zea mays subsp. mays	United States	Nebraska
PI	594071	N523	Zea mays subsp. mays	United States	Nebraska
PI	594072	N524	Zea mays subsp. mays	United States	Nebraska
PI	594073	N525	Zea mays subsp. mays	United States	Nebraska
PI	594074	N526	Zea mays subsp. mays	United States	Nebraska
PI	594075	N528	Zea mays subsp. mays	United States	Nebraska
PI	594076	N529	Zea mays subsp. mays	United States	Nebraska
PI	594077	N530	Zea mays subsp. mays	United States	Nebraska
PI	594078	N532	Zea mays subsp. mays	United States	Nebraska
PI	594079	N533	Zea mays subsp. mays	United States	Nebraska
PI	594080	N534	Zea mays subsp. mays	United States	Nebraska
PI	594081	N535	Zea mays subsp. mays	United States	Nebraska
PI	594084	N538	Zea mays subsp. mays	United States	Nebraska
PI	594087	N541	Zea mays subsp. mays	United States	Nebraska
PI	594088	N542	Zea mays subsp. mays	United States	Nebraska
PI	594089	N543	Zea mays subsp. mays	United States	Nebraska
PI	594090	N544	Zea mays subsp. mays	United States	Nebraska
PI	595366	N209	Zea mays subsp. mays	United States	Nebraska
PI	595541	CML 247	Zea mays subsp. mays	Mexico	Federal Dis.
PI	596354	N211	Zea mays subsp. mays	United States	Nebraska
PI	596355	N216	Zea mays subsp. mays	United States	Nebraska
PI	596357	N218	Zea mays subsp. mays	United States	Nebraska
PI	597578	N546	Zea mays subsp. mays	United States	Nebraska
PI	600755	LP1 CMS HT	Zea mays subsp. mays	United States	
PI	600772	FR19	Zea mays subsp. mays	United States	Illinois
PI	600944	LH39	Zea mays subsp. mays	United States	Iowa
PI	600957	LH74	Zea mays subsp. mays	United States	Iowa
PI	600958	FAPW	Zea mays subsp. mays	United States	
PI	601008	PHG35	Zea mays subsp. mays	United States	Iowa

Supplementary Table 2 continued					
PI	601009	B47	Zea mays subsp. mays	United States	Iowa
PI	601037	G80	Zea mays subsp. mays	United States	Iowa
PI	601079	LH123HT	Zea mays subsp. mays	United States	Iowa
PI	601210	78004	Zea mays subsp. mays	United States	
PI	601301	78002A	Zea mays subsp. mays	United States	
PI	601319	PHG72	Zea mays subsp. mays	United States	Iowa
PI	601320	PHG84	Zea mays subsp. mays	United States	Iowa
PI	601322	PHZ51	Zea mays subsp. mays	United States	Iowa
PI	601403	LH156	Zea mays subsp. mays	United States	
PI	601438	78371A	Zea mays subsp. mays	United States	
PI	601441	PB80	Zea mays subsp. mays	United States	
PI	601466	LH59	Zea mays subsp. mays	United States	
PI	601468	PHK29	Zea mays subsp. mays	United States	Iowa
PI	601489	740	Zea mays subsp. mays	United States	Minnesota
PI	601493	LH149	Zea mays subsp. mays	United States	
PI	601494	LH65	Zea mays subsp. mays	United States	
PI	601499	PHT77	Zea mays subsp. mays	United States	Iowa
PI	601500	PHV63	Zea mays subsp. mays	United States	Iowa
PI	601501	PHW65	Zea mays subsp. mays	United States	Iowa
PI	601561	6M502	Zea mays subsp. mays	United States	
PI	601574	PHT60	Zea mays subsp. mays	United States	Iowa
PI	601610	H8431	Zea mays subsp. mays	United States	Minnesota
PI	601684	WIL900	Zea mays subsp. mays	United States	
PI	601685	WIL901	Zea mays subsp. mays	United States	
PI	601686	WIL903	Zea mays subsp. mays	United States	
PI	601725	J8606	Zea mays subsp. mays	United States	Minnesota
PI	601726	L 127	Zea mays subsp. mays	United States	
PI	601728	L 139	Zea mays subsp. mays	United States	
PI	601729	W8555	Zea mays subsp. mays	United States	Minnesota
PI	601777	PHK35	Zea mays subsp. mays	United States	Iowa
PI	601778	PHM10	Zea mays subsp. mays	United States	Iowa
PI	601782	PHN73	Zea mays subsp. mays	United States	Iowa
PI	601784	PHP55	Zea mays subsp. mays	United States	Iowa
PI	638550	N552	Zea mays subsp. mays	United States	Nebraska
PI	601788	PHT22	Zea mays subsp. mays	United States	Iowa
PI	601789	PHV37	Zea mays subsp. mays	United States	Iowa
PI	604606	N527	Zea mays subsp. mays	United States	Nebraska
PI	606329	DE1	Zea mays subsp. mays	United States	Delaware
PI	606768	SD40	Zea mays subsp. mays	United States	South Dakota
PI	606769	SD41	Zea mays subsp. mays	United States	South Dakota

Supplementary Table 2 continued					
PI	607512	N7A	Zea mays subsp. mays	United States	Nebraska
PI	633840	Tx714	Zea mays subsp. mays	United States	Texas

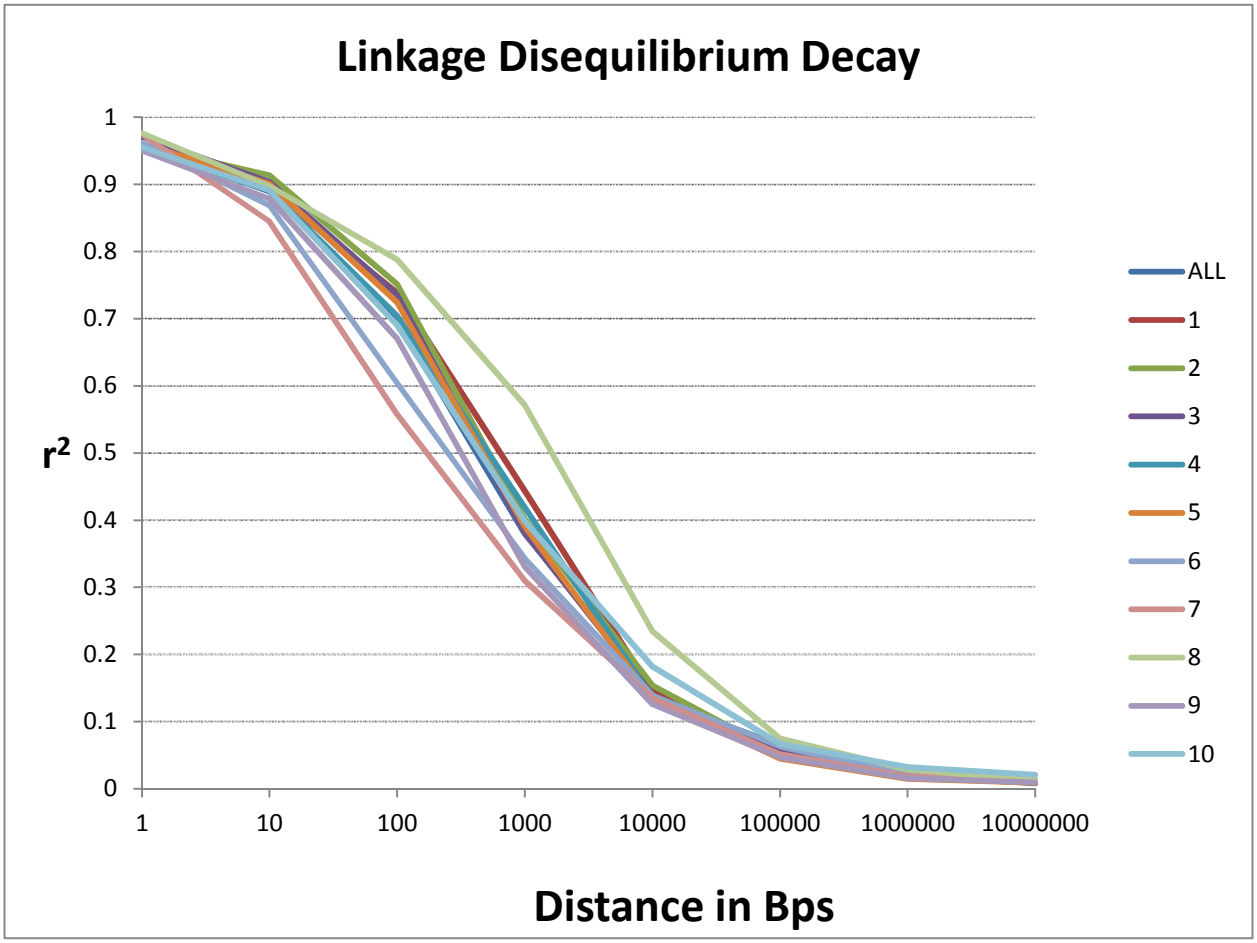


Figure 1. Linkage disequilibrium decay measuring r^2 over base pairs across all 10 maize chromosomes.

Figure 2. Population structure estimates based on 1665 SNPs distributed across the maize genome. The area of 2 different colors (Red and Green) illustrates the proportion of each subpopulation based on these SNPs markers.

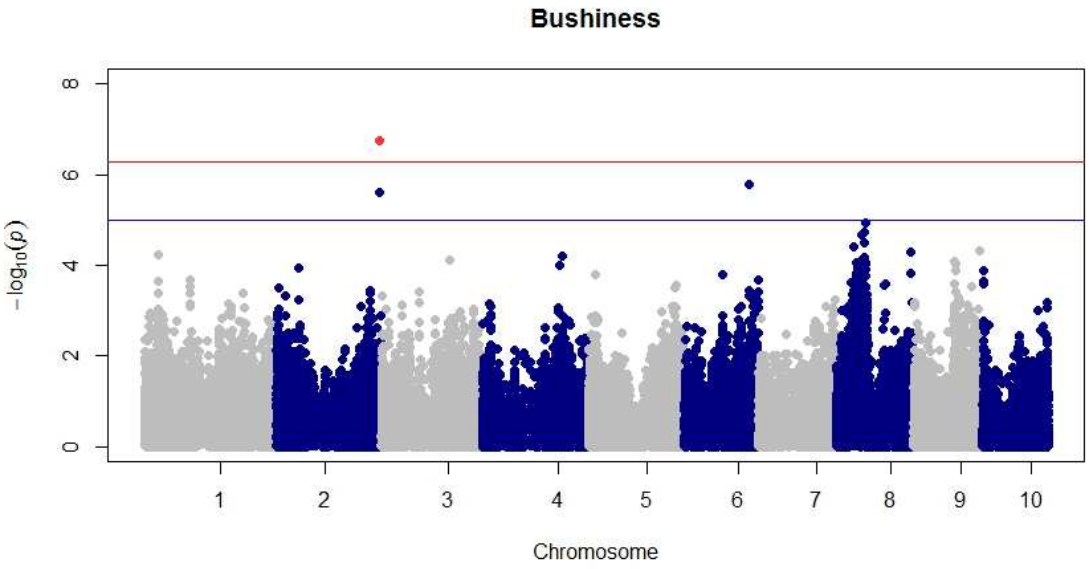


Figure 3. Manhattan plot showing associations between individual polymorphisms through the entire maize genome for **BSH.** MLM was used fitting both Q and K matrix. Only one marker on chromosome 2 was found to be significant at $p<5.23 \times 10^{-7}$

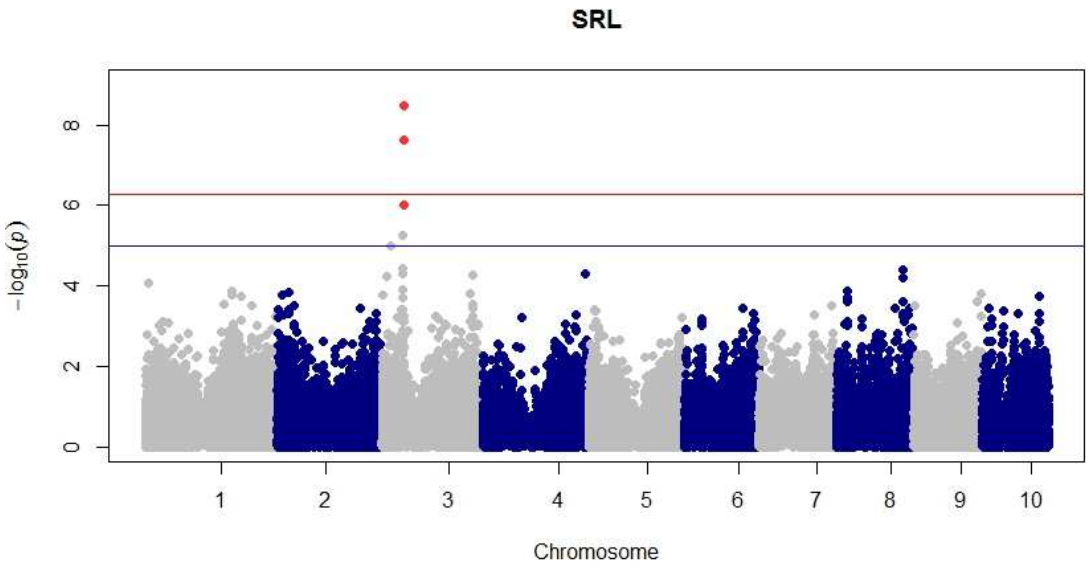
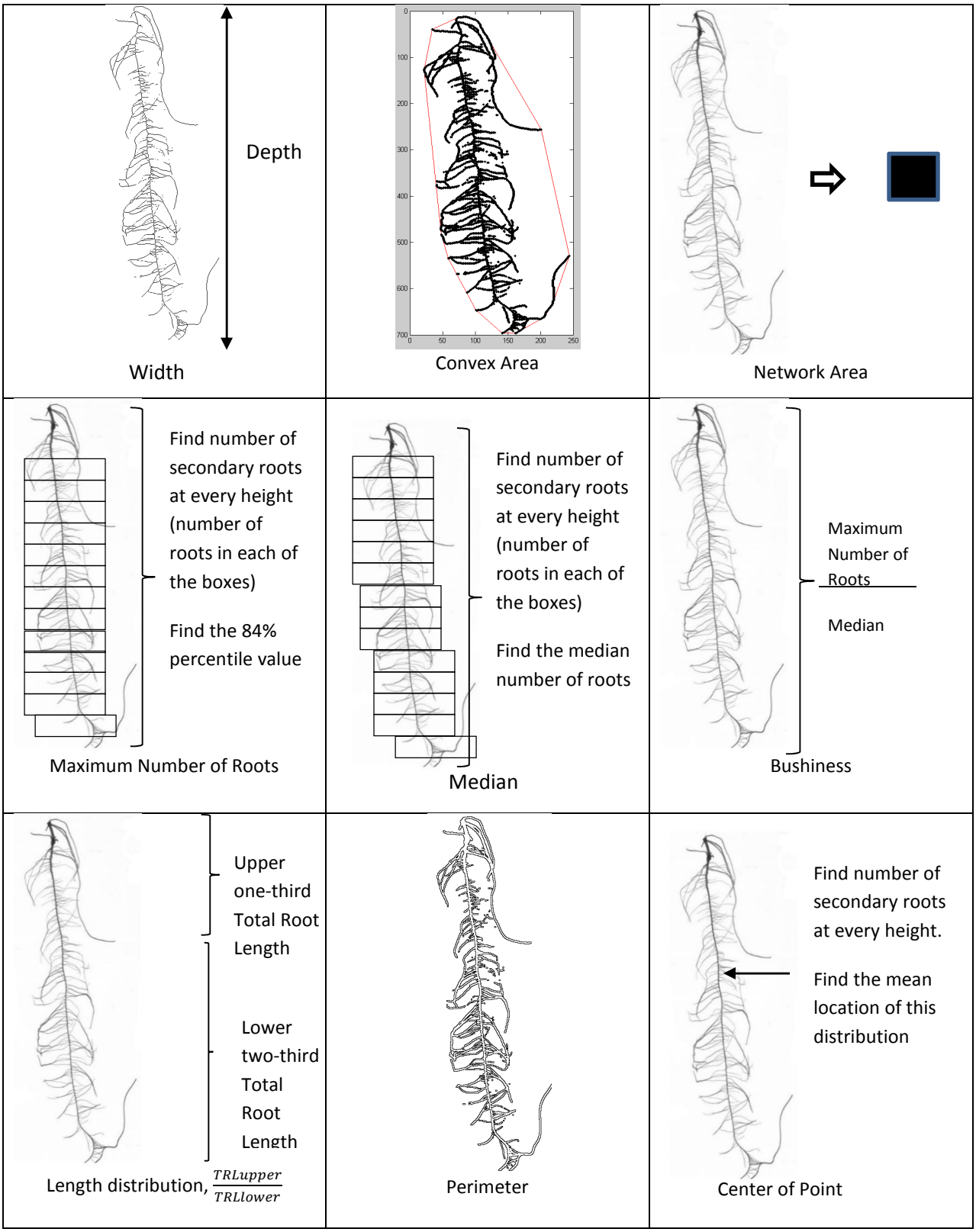
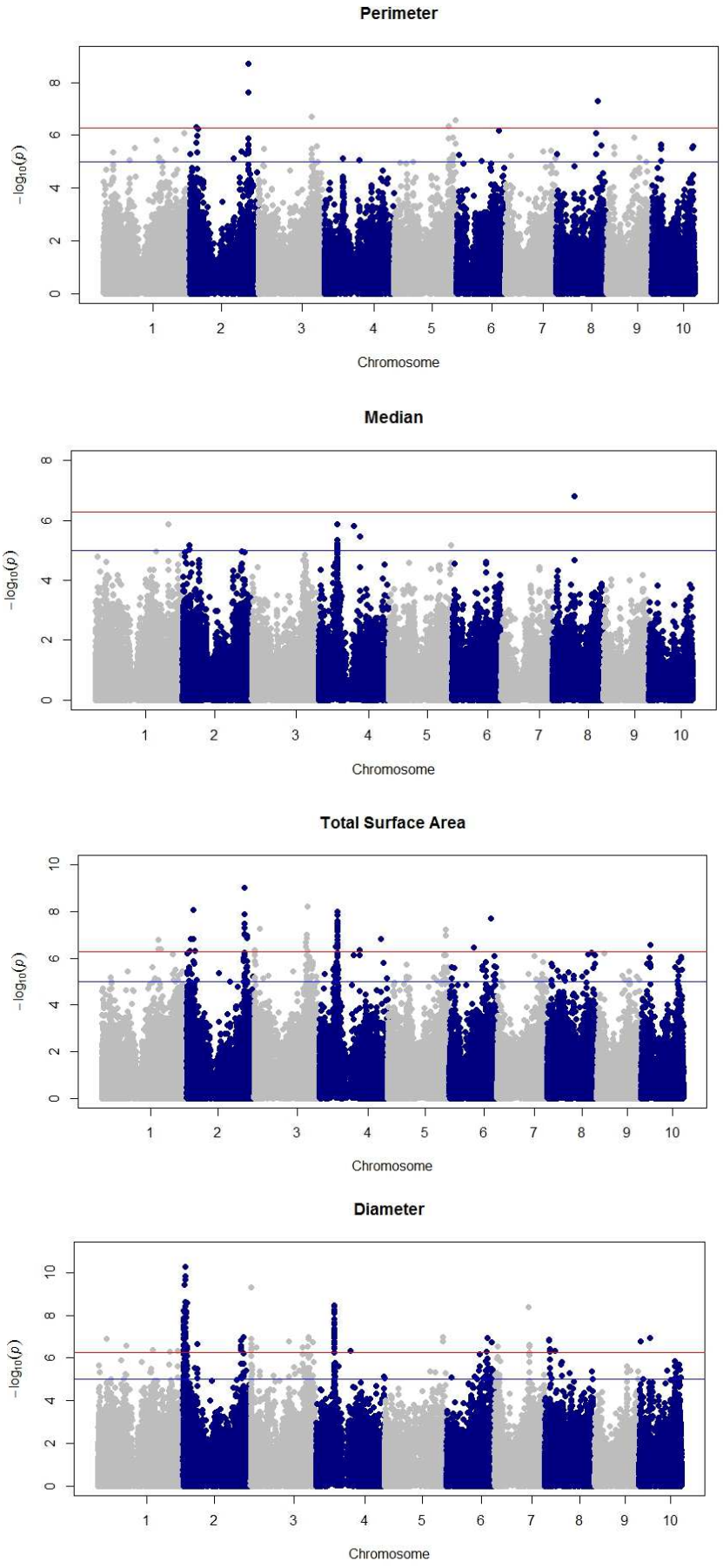


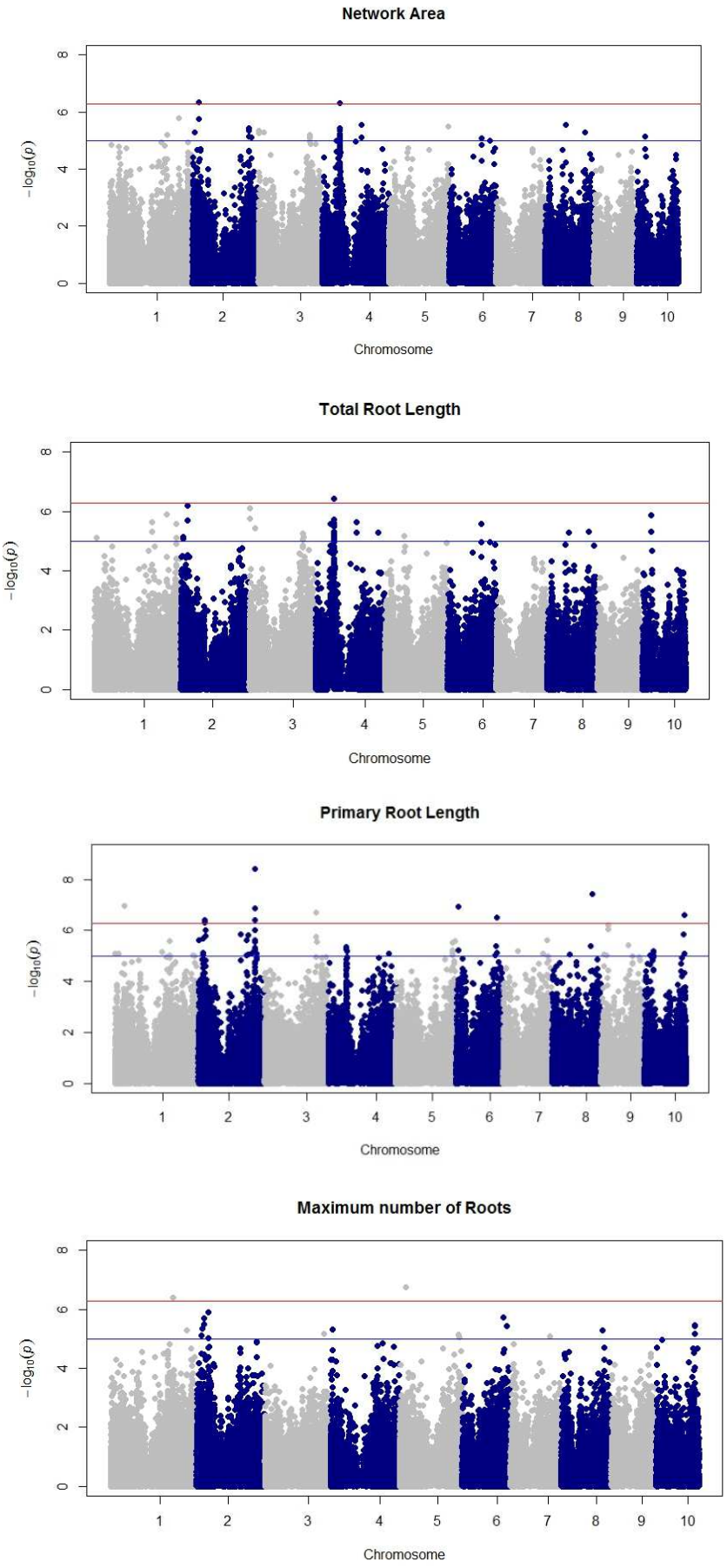
Figure 4. Manhattan plot of GWAS using MLM. Marker trait associations with SRL are shown across the entire genome. Peaks are found on chromosome 3 only using a threshold of $p<5.23 \times 10^{-7}$



Supplementary Figure 1. Illustrations of the parameters measured by *ARIA* for seedling root traits extracted for GWAS.



Supplementary Figure 2. Genome-wide Manhattan plots of significant loci using GLM.



Supplementary Figure 2 continued. Genome-wide Manhattan plots of significant loci using GLM.

CHAPTER FIVE

GENOMIC PREDICTION OF SEEDLIGN ROOT LENGTH IN MAIZE (*ZEA MAYS* L.)

Jordon Pace[1*], Xiaoqing Yu[1], and Thomas Lübberstedt[1]

Paper is in preparation for submission to the Plant Journal. Abstract, structure, and references are all formatted according to journal standards.

Summary:

Genotypes with extreme phenotypes are valuable for studying "difficult" quantitative traits. Genomic prediction (GP) might allow identifying such extremes, by phenotyping a training population of limited size and predicting genotypes with extreme phenotypes in large sequences germplasm collections. We tested this approach employing seedling root traits in maize and the extensively genotyped Ames Panel. A training population made up of 384 inbred lines from the Ames Panel was phenotyped by extracting root traits from images using the software program *ARIA*. A Ridge Regression BLUP (RR-BLUP) strategy was used to train a GP model. Genomic Estimated Breeding Values (GEBVs) for the trait Total Root Length (TRL) were predicted for 2,431 inbred lines, which had previously been genotyped by sequencing. Selections were made for TRL 100 extreme lines each with predicted longest or shortest TRL were validated for TRL and other root traits. The two predicted extreme groups with regard to TRL were significantly different ($p=0.0001$). The difference of predicted means for TRL between groups was 145.1 cm, and 118.7 cm for observed means, which were significantly different ($p=0.001$). The accuracy of predicting the rank 1-200 of the validation population based on TRL, longest to shortest was determined using a Spearman correlation to be $\rho=0.55$. Taken together, our results support that

GP may be a useful approach to identify the most informative genotypes in sequenced germplasm collections to facilitate experiments for quantitative inherited traits.

Introduction

The ability to predict superior genotypes with high accuracy is of key importance in plant breeding. Marker assisted selection (MAS) has become a routine procedure in commercial breeding programs because of an increased gain per unit time when compared to phenotypic selection (Eathington et al. 2007). Genomic Prediction (GP), a form of MAS, has become a valuable tool in animal breeding and recently been shown to be reliable in crop breeding as well (Sallam et al. 2014). GP is not only a promising approach for breeding purposes, but also for basic research. GP enables to identify extreme genotypes for traits that are difficult to measure for large numbers of individuals. In this way the most informative genotypes for mapping or expression studies can be identified in large sequenced genotype collections, which increasingly become available for crop species. The Ames Panel in maize (Romay *et al.* 2013) for example, consists of 2815 inbred lines with genotyping by sequencing (GBS) based single nucleotide polymorphism (SNP) data readily available. Other collections of genetic resources in maize include the nested association mapping population (NAM) (Yu *et al.* 2008), the maize association mapping panel (http://www.panzea.org) (Yu *et al.* 2006), and the IBM population (Lee *et al.* 2002) (MaizeDB, http://www.agron.missouri.edu/), all three with readily available genotypic and phenotypic data.

When performing quantitative trait locus (QTL) or association mapping, selecting parents or individuals used for the mapping population is critically important. Use of lines with extreme phenotypes will ensure that the population is segregating for QTL controlling a trait of interest

(Mackay and Powell 2007). Identifying extreme genotypes is the basis of bulked segregant analysis (BSA), originally used to identify markers for disease resistance genes in crops in which genotypes with contrasting phenotypes are pooled in separate groups to identify markers associated with those traits (Michelmore *et al.* 1991). Using GP, subsets of large diversity panels can be used to collect phenotype data and train a prediction model for traits of interest while exploiting markers covering the genome at high density. GP could then be performed on the complete panel to identify genotypes with likely extreme phenotypes. This strategy would decrease the amount of resources needed for often laborious and costly phenotyping as only the training population requires intense phenotyping prior to validation (Meuwissen *et al.* 2001). This would facilitate studying "difficult" traits with the genetically most informative genotypes.

There are numerous GP approaches, notably Bayesian and mixed model procedures. No single method is superior in all circumstances (Bernardo and Yu 2007, Lorenzana and Bernardo 2009, Heslot *et al.* 2012). Ridge regression best linear unbiased prediction (RR-BLUP) (Meuwissen *et al*. 2001) has been shown to be a reliable model in providing good prediction accuracy for a range of quantitative traits in crop species (Heslot *et al.* 2012). The mixed model used by RR-BLUP shrinks all marker effects equally with the same variance for a trait. This is in accordance with the infinitesimal model of quantitative genetics and explains why RR-BLUP performs well for complex trait predictions (Clark *et al.* 2011). Computationally, RR-BLUP is less demanding than Bayes A, B, and Cπ, making it one of the more approachable and suggested methods of genomic prediction (Heslot *et al.* 2012). GP is conducted through a three-step process. First, a training panel of a representative sample of individual lines is developed in order to obtain estimates for marker effects. Second, this marker information can be used to determine genomic estimated breeding values (GEBVs) for any genotype with available marker

information. Finally, genotypes are selected based on GEBVs in a breeding population (Meuwissen 2001). When performing any form of GP, the selection of the training panel is of key importance as it should be a good representation of the breeding population (Heffner *et al.* 2009). For populations that are many generations apart, or quite different in composition, an increased number of genotypes are needed to increase prediction accuracy (Rutkoski 2010). Another challenge with GP is accuracy of phenotyping. High accuracy concerning phenotype measurements in the training panel is critical for prediction of GEBVs based solely on genomic information. Prediction accuracies are affected by the amount of linkage disequilibrium (LD) within the training population as compared to the validation population, genetic relationships between the validation and training population, the genetic architecture and heritability of target traits, marker density, and training population size (Hayes *et al.* 2009, Luan *et al*. 2009, Zhong *et al*. 2009).

Difficult to measure quantitative traits that would benefit from the use of GP in maize or any crop species are root architectural traits. The maize root system is an integral part of plant growth and productivity (Lynch, 1995, Aiken and Smucker 1996). Root architecture plays a major role in plant nitrogen use efficiency (NUE) as well as drought tolerance (Ribaut *et al.* 2007). The large variability in root architecture is an unexploited opportunity to select for beneficial root architectural traits that increases crop production in nutrient and moisture deficient environments (Lynch et al. 2014). Root traits have not been used for selection as they show generally low heritability, and there are no accurate, fast measurements, allowing high-throughput field measurements comparable to determining grain yield using high-throughput harvesting equipment such as combines (Malamy 2005, Tuberosa 2012). One method to alleviate some of these issues is through the development of software that can extract multiple

root traits from a single image (Pace *et al*. 2014) enabling a more high-throughput approach to root architecture characterization. Another currently available trait collection method involves collection of mature root systems in field trials using both mobile reproducible image acquisition and later algorithm based trait extraction (Bucksch *et al.* 2014).

Recent studies have shown that phenotyping roots at a seedling stage in hydroponic conditions could help to alleviate some of the time and resources required for large scale root studies (Kumar *et al.* 2014, Pace *et al.* 2015, and Pace *et al.* 2014). Genomic prediction for root traits at the seedling stage has not yet been tested in order to identify extreme genotypes or to determine prediction accuracy. The objectives of this study were to 1) determine accuracy of genomic prediction based on three simple root traits; 2) determine genomic prediction's ability to accurately rank genotypes for root traits and identify genotypes with extreme root lengths; 3) determine the effects of combining the training population and validation population on cross-validation accuracy estimates; and 4) validate a previous GWAS study within a larger population for root traits collected by *ARIA*.

Results

Training population cross-validation

Phenotypic measurements of the trait TRL, along with genotypic information encompassing 186,849 markers across the maize genome of the 384 line training population were used along with RR-BLUP to train the prediction model. Ridge Regression BLUP (RR-BLUP) was chosen because it is less computationally intensive and it resulted in high accuracies in predicting quantitative traits with multiple small effect QTL such as TRL (Heslot *et al.* 2012). In order to determine, if TRL was a suitable trait for GP, cross-validation was completed with a random

60/40 training/validation population split within the 384 inbred line population. Average accuracy was calculated after running RR-BLUP with 500 iterations. Within each iteration, randomly selected lines were used as training or validation population, respectively accuracies ranged from r=0.10-0.56 with an average accuracy of r=0.42 (Figure 1). Two additional root length based traits were used for cross validation, Primary Root Length (PRL) and Secondary Root Length (SEL). SEL is closely correlated with TRL (r=0.98), while PRL showed a lower correlation (r=0.70) (Pace *et al.* 2015). Using the same strategy as for TRL, iterating cross-validations 500 times for PRL gave similar accuracy ranges as for TRL, ranging from r=0.10-0.58, with an average accuracy of r=0.44 (Figure 1). For SEL one cross-validation run gave a similar accuracy of r=0.45. When running RR-BLUP with 500 iterations, the range and standard deviation of accuracies decreased to r=0.30-0.56, with an average accuracy of r=0.43 (Figure 1). All three root length based traits had similar ranges and average accuracy estimates from cross-validation within the training population. These accuracy values gave confidence that TRL was suitable in giving sufficiently accurate predictions.

Group and ranking prediction accuracy

The same marker set was used to predict root length traits in the validation population. The validation population consisted of the remaining 2,431 maize inbred lines of the Ames Panel (Romay *et al.* 2013). Based on estimated marker effects, GEBVs of TRL on all 2,431 remaining inbred lines were predicted. Selections for empirical validation were made by selecting the 100 genotypes with the highest TRL estimates and the 100 genotypes with the lowest TRL estimates. These 200 genotypes were tested in replicated experiments under the same growing conditions as the training population. These selections were based on TRL alone and were not the predicted extremes for PRL and SEL.

To determine whether GP accurately predicted which inbred lines had long or short root systems, a group effect was added to the linear model in order to analyze the difference in means for each group. Lsmeans were calculated for each subset with predicted long (LONG) or short root systems (SHORT). LONG had an observed TRL Lsmean of 295.1 cm, larger than the predicted average GEBV for LONG of 246.1 cm. SHORT, had an observed TRL Lsmean of 176.4 cm (Table 1), also larger than the predicted GEBV for TRL at 100.8 cm. The observed difference for average TRL between LONG and SHORT was 118.7 cm while the predicted difference of averages was 145.3 cm. To determine, if the difference between the LONG and SHORT groups were accurately captured by GP, pairwise comparisons between LONG and SHORT were made for both the predicted and observed values. Individuals were ranked longest to shortest within each group, the difference between corresponding ranks 1-100 (Highest ranked predicted genotype TRL in LONG subtracted by the highest predicted genotype in SHORT), were calculated for both groups. The Lsmean of those differences between groups was than calculated for both predicted and observed values. Group differences (predicted difference Lsmean = 144.7 and observed difference Lsmean = 120.2) were found significantly different ($p = 0.001$). This significant difference between extreme group differences between the groups we observed vs. predictions indicates that although GP could accurately identify extreme genotypes, the predicted difference between the two groups was not equal to the difference observed. This difference between group means for the predicted mean GEBVs and observed Lsmeans is likely due to the overall underestimation of TRL seen throughout the validation population and can be attributed to the low accuracy within each group (LONG: $r=0.12$, SHORT: $r=0.10$). These observed group values Lsmeans for TRL were found to be significantly different ($p=0.0001$). Furthermore, the ranking of each genotype's predicted and observed TRL values were compared

to determine our ability of accurately predicting the rank of selected genotypes by root length. The Spearman ranking correlation between predicted and observed ranks for the 200 genotypes was ρ=0.55 and significant at p=0.0001 (Figure 2). Ranking accuracy within each 100 line validation group was much lower for TRL ρ=0.10 for LONG and below ρ=0.05 for SHORT. For comparison, PRL and SEL were also predicted upon within the same 200 line validation population. The spearman ranking correlation for SEL was very similar to TRL at ρ=0.54 while PRL was lower at ρ=0.38 (Table 2). The same trend of lower accuracies within each group was observed for both PRL and SEL with PRL having the highest within group accuracy within the LONG group at ρ=0.29.

Genomic prediction

Predicted TRL GEBVs for the validation population ranged from a minimum of 72.6 cm to a maximum length of 286.7 cm. Observed TRL Lsmean values for the validation population had a much wider range from 37.1 cm minimum to a maximum observed of 532.3 cm. Taken as a whole, the Pearson correlation between predicted TRL of the validation population and observed TRL for the validation population showed an accuracy of r=0.59 (Figure 3). To determine within group accuracies, Pearson correlations were calculated within each of the two subgroups. Accuracies were much lower with r=0.12 and r=0.10 for LONG and SHORT, respectively.

To test the effect of adding the validation population on the accuracy of genomic prediction for TRL, the observed Lsmeans for the 200 lines selected for the validation population were added to those of the original training population to create a population consisting of 584 inbred lines. As before, a 60/40 training /validation population cross-validation test was performed with 500 iterations. The range of accuracies was much narrower than within the

original 384 training population. Accuracy estimates ranged from r=0.40 to r=0.60 with an average accuracy at r=0.54, reducing the amount of standard error. For the two other length based traits, PRL and SEL, accuracies decreased compared to the 384 line population to r=0.31 for PRL and r=0.32 for SEL (Table 3 and Figure 1).

Accuracies across subpopulations

The original 384 line training population was split into two subpopulations based on STRUCTURE 2.3.4 analysis from previous work (Pace *et al.* 2015). The two subpopulations varied in size, the larger subpopulation consists of 319 and the smaller subpopulation of 65 lines. The larger subpopulation contains mainly non-stiff stalk and mixed group lines, the smaller subpopulation consists of mostly stiff-stalk lines. When using the large subpopulation as a training population, the prediction accuracy for TRL was r=0.45. In contrast, when using the small subpopulation as a training population, accuracies were lower at r=0.29. When completing 60/40 cross validations within each sub population, the large subpopulation average accuracy was r=0.31 while the small subpopulation cross validation mean accuracy was r=0.39. When using STRUCTURE 2.3.4 for the 584 line population three subpopulations were identified, Q1, Q2, and Q3 (Figure 4), comprising of 83, 390, and 103 lines, respectively, with 8 lines highly mixed between all three subpopulations. Using Q2 as training population to predict performance of Q3, a prediction accuracy of r=0.37 was obtained. With Q3 as training set, a lower prediction accuracy of r=0.21 was estimated when Q2 was the validation population. Prediction accuracy estimates involving Q1 either as training or validation population was negative with both Q2 and Q3.

GWAS validation

GWAS validation of a previous study (Pace *et al.* 2015) was completed using the new 584 population for all traits. A mixed linear model (MLM) and general linear model (GLM) GWAS were performed for all 22 corresponding root traits in the previous study. No markers were found significant for TRL, SEL, or PRL using MLM, but the marker with the lowest P-value for both TRL and SEL was the same (S5_175865830). All three traits showed significant marker associations using GLM, TRL with 5, SEL with 8, and PRL with 129 significant markers. All three displayed more significant markers using GLM than compared to the previous study (Pace *et al.* 2015) where TRL had one significant marker, SEL had zero significant markers, and PRL had 11 significant markers. For SEL and TRL, marker S5_17865830 was significant and found in an intergenic region (B73 RefGen_V2). Marker S3_223308733 was significant for TRL, SEL, and PRL at the significance cutoff of $p=5.3 \times 10^{-7}$. This marker was found within gene model GRMZM2G336017, which codes for a hypothetical protein with no known function or expression data currently.

Discussion

The purpose of this study was to determine whether GP enables identification of extreme genotypes using a subset of lines from a larger sequenced population, using maize seedling TRL as model trait. Genetic resource collections with readily available genotype data are available for maize; other important crops with available genetic resources include sorghum, soybean, barley, rice and wheat (Liang *et al.* 2007, Morris *et al*. 2012, Munoz-Amatriain *et al.* 2014. Spindel *et al.* 2015, and B. Diers, personal communication). Decreasing costs in genotyping will stimulate development of additional sequenced genetic resources in the near future. Thus, a similar GP

strategy as presented herein will be applicable to an increasing number of species with respective resource collections. Without the need of genotyping, GP would expedite studies of quantitative traits. GP reduces phenotyping efforts and identifies the most informative genotypes. Smaller subsets of lines, perhaps core collections of gene banks, could be characterized in detail for traits of interest to train prediction models and predict extreme genotypes within large genetic resource collections. Those selected individuals are more likely to carry rare alleles that affect traits of interest, which, due to low allele frequencies, would otherwise be underrepresented in random inbred line panels typically used for GWAS studies.

Identifying extreme genotypes

We evaluated, whether GP would capture significant difference between the LONG and SHORT groups selected solely based on predictions of GEBV for TRL within a large validation population (2431 inbred lines). The two groups LONG and SHORT, were significantly different for TRL (p=0.0001). Thus, GP successfully predicted length of seedling roots with satisfactory accuracy ($r_{\hat{y},y}$=0.59 and $r_{\hat{g},g}$= 0.91) when using a training population of modest size (384 inbred lines). It should be noted, however, that a total of 24 LONG inbred lines had a shorter observed TRL than SHORT lines. Also, neither the longest predicted nor the shortest predicted line were actually the lines with longest or shortest TRL within their respective groups. However, both were assigned correctly to LONG and SHORT, respectively. Predicted GEBVs for TRL were underestimated when compared to observed values of TRL for both groups. To evaluate this, we regressed phenotypes (y) onto GEBVs ($\hat{y}$) ($\beta_{y\hat{y}}$=0.82). This analysis showed that for every unit change of our observations we only had 82% of that change explained by our predictions resulting in the underestimated TRL predictions compared to observations. The LONG group's

observed average was 49.1 cm longer than predicted. For the SHORT group the observed average TRL was 75.5 cm longer than predicted. This difference and ultimately underestimation of TRL using RR-BLUP is likely due to the even shrinkage effect on all markers throughout the genome (Endelman 2011). Overall, the ability of GP to capture differences between two extreme groups for a moderately heritable trait is encouraging even though ranking was not perfectly predicted. This approach will still reduce overall phenotyping efforts and offers a solution to effectively identify informative individuals for mapping and other studies.

For plant breeding, the ability to predict ranking of lines is important to save resources for further evaluation. Spearman rank correlations of predicted and observed values for TRL were significant with moderate high accuracy of ρ=0.55 and adjusted prediction accuracy of ρ=0.84 (Table 3). Within each group the ranking accuracy was 0.1 for LONG and even lower for SHORT. This within grouping ranking should be considered when determining the number of individuals selected for each extreme group as the ranking within groups will likely differ from predictions (Daetwyler *et al*. 2013).

Traits PRL and SEL were evaluated for ranking accuracy. SEL showed a similar ranking accuracy (ρ=0.54) as TRL, while that of PRL was lower (ρ=0.38) (Table 2). Again within group ranking accuracies were much lower for both SEL and PRL for both LONG and SHORT (ρ=0.04 and 0.15, respectively). Accuracies of the validation population may be inflated due to the fact that these lines were not selected at random but based on predicted extremes for long and short roots.

Comparison of accuracies

RR-BLUP was shown to be a suitable method for predicting root architectural traits as it pertains to the length of roots in a controlled environmental setting. Accuracies based purely on predicted estimates of TRL correlated to observed measurements ($r_{\hat{y},y}$) ranged from r=0.21-0.54. When looking at the adjusted accuracy ($r_{g\hat{},g}$) the ranges were higher at r=0.33-0.83. The higher accuracies for TRL follow the trend described in previous prediction studies showing that an increase in population size generally increases prediction accuracies, in particular because more individuals are used for training the prediction model (Asoro *et al*. 2011, Zhong *et al*. 2009). By selecting the predicted extremes, one would expect that adding these groups to the initial training population would increase the correlation between predicted and observed phenotypes.

As selections were based solely on TRL, extreme lines for TRL are not necessarily extreme for PRL and SEL. When performing cross-validation for SEL and PRL, both followed the exact opposite trend when adding lines to the total population. In the initial 384 line population, accuracies were r=0.43 and r=0.44 for SEL and PRL, respectively. After adding the additional 200 lines from the validation population, cross-validation accuracies decreased to 0.32 and r=0.31, respectively (Table 3). This decrease is likely due to the fact that the genotypes were selected based on TRL alone, these genotypes are likely not all extremes for PRL and SEL. While accuracies decreased for SEL and PRL after adding the validation population, the standard deviation of the 500 estimates of accuracy for each cross-validation decreased for all three root traits (Table 3). These results show that increasing the number of individuals within the training population can reduce error and give more precise accuracy estimates.

We conducted a comparison of accuracies based on different population and subpopulation compositions of training and validation populations. In general, prediction accuracies were reduced across subpopulations compared to using random sets of lines for cross-validation. In case of the 384 line population, one subpopulation is composed mainly of non-stiff stalk (NSS) and mixed lines while the second smaller subpopulation is composed of mostly lines from the stiff stalk (SS) heterotic group. A decrease in prediction accuracy using the NSS subpopulation to train the model as compared to using the SS subpopulation is confounded by a decrease in relatedness between the training and validation population, and also the large difference in number of individuals used to train the prediction model. These two confounding factors are explained by (i) poor predictions between heterotic groups due to limited relatedness between populations and (ii) population size. One subpopulation consists of 319 lines and the other of only 65 lines. This large discrepancy is an example of how the number of individuals used to train and predict affects prediction accuracy. For the 584 line population, the subpopulation stratification added a third subpopulation. Two subpopulations, Q2 and Q3 (Figure 4), contained similar lines as the original 384 line population with many SS and NSS stalk lines within both. The third subpopulation, Q1 (Figure 4), is quite different with many lines with no heterotic group designations found within (Romay *et al.* 2013) as well as the Germplasm Resources Information Network (GRIN). Predictions made using Q2 or Q3 on Q1 gave negative correlations with poor accuracy. Many of the lines within Q1 do not share pedigree information with lines from the other subgroups and, therefore, represent an example of how relationships between the training and validation populations can affect prediction accuracies, including negative effects. Asoro *et al.* (2011) also found decreased prediction accuracies in oats, when training and validation populations were not related. As the relatedness between training and

validation population decreases, increasing the size of the training population can sometimes offset a reduction in accuracy (Rutkoski 2010).

GWAS validation

The genomic prediction method RR-BLUP uses the infinitesimal model of quantitative genetics and assumes normally distributed effects across the whole genome (Meuwissen 2001). Studies in animal breeding have shown that using GP models such as Baysian models or G-BLUP also reduce the risk of identifying false positives when trying to detect QTL (Zeng *et al.* 2012). GP likely increased the power to detect rare alleles affecting traits of interest. As genotypes that exhibit extreme phenotypes in both directions with rare alleles are likely hard to detect using GWAS based on randomly selected genotypes. Using this GP strategy, one could directly move into mapping studies such as BSA, which can be used for both qualitative and quantitative traits (Liu *et al.* 2012 and Venuprasad *et al.* 2009). One could also select a portion of the individuals found within each extreme group that could be crossed and used to develop a bi-parental population for linkage mapping, or directly use these extreme genotypes for association mapping. Ultimately, GWAS has not identified many reliable SNPs and those identified still must be validated before used for developing improved germplasm through processes such as MAS, marker assisted recurrent selection (MARS), or F_2 enrichment [$R_{eference}$]. In contrast, GP has given moderate to high prediction accuracies and ranking accuracies for TRL, SEL, and PRL when all markers are considered simultaneously. This shows that for purposes of germplasm development, GP is more useful for complex traits when compared to GWAS. While ignoring underlying genetic and molecular mechanisms, GP is still able to capture much more of the variation than GWAS and therefore a quicker and more reliable tool for breeding purposes.

There have been several studies on genetic control of root architecture under various growing conditions (Hund et al. 2011). Many of these studies found few or no overlapping QTL for root development traits. Discrepancies are likely caused by population parameters such as linkage disequilibrium, allele frequencies found between mapping populations and heritability of the traits studied (Yu *et al*. 2006). As mapping studies using different mapping populations studying complex traits such as root architecture do not often find many overlapping significant loci or candidate genes (Bernardo, 2008), new strategies are needed to help identify causative rare alleles. Another obstacle in validating highly quantitative traits is a lack of statistical power to detect low effect QTL associated with traits of interest. This problem has also been shown in a lack of continuity between multiple root QTL and association studies reported (Hund *et al*. 2009, Kumar *et al.* 2014, Pace *et al.* 2015, Pace *et al*. 2014). A validation experiment for the same traits under the same growing conditions did not find many consistent loci (Pace *et al.* 2015). This is consistent with earlier reports on "missing heritability" in GWAS studies (Maher 2008), i.e. most of the detected loci explain very little of the heritability for a given trait and we are not able to detect rare variants. Including extremes identified using GP can enrich the population with rare alleles that effect traits of interest at a frequency more easily detectable than with a random admixed population.

The 584 line GWAS population created using GP within this study is likely more relevant not only due to an increase in size, but also because of enrichment for alleles impacting root trait TRL. Even though no true validations were made, there was an increase in the number of significant markers identified for TRL, SEL, and PRL compared to Pace *et al.* (2015). Furthermore a common marker S5_17865830 was found to have the lowest p-value within MLM and it was significant under GLM for TRL and SEL. Marker S3_223308733 was found

significant for TRL, SEL, and PRL. This marker in particular is found within or close by (<10 kb) three gene models with possible function in root development at the seedling stage. Gene model GRMZM2G336017 codes for a hypothetical protein and there are currently no expression data found within the maize expression atlas (Sekhon *et al.* 2011). There are two other gene models for which marker S3_223308733 is in LD. Gene model GRMZM2G034943 is found upstream of the marker at location 223,308,255 with a second gene model GRMZM2G035134 located downstream of the marker at location 223,309,446 according to B73 RefGen_v3. Both gene models have the GO term primary root and moderate expression identified by Sekhon *et al.* (2011). These two genes may offer good candidates as being involved in the development of the primary root especially as a significant association was found for trait PRL. The identification of possible root development candidate genes and overall increase in number of markers found significant using GLM for all three traits supports that power to detect putative QTL for root traits was increased when compared to the previous study (Pace *et al.* 2015) employing a smaller population size for GWAS.

Experimental Procedures

Plant materials

Our study is based on 584 inbred lines from the Ames Panel (Romay *et al*. 2013) acquired from the USDA-ARS North Central Regional Plant Introduction Station (NCRPIS) in Ames, Iowa. The training population is a subset of 384 lines, which have been used in a Genome Wide Association Study (GWAS) previously (Pace *et al*. 2015). The remaining 200 inbred lines were selected from the remaining about 2400 lines in the Ames Panel based on predictions

calculated using RR-BLUP for total root length (TRL): 100 lines each with the predicted longest and shortest root lengths were selected as validation panel.

Root phenotyping

The 200 line validation population was grown under the same growing conditions and for the same duration of time as the training population (Pace *et al*. 2015). Briefly, seeds were sterilized with Clorox solution (6% sodium hypochlorite) for 15 minutes, and washed twice thereafter using autoclaved water. Germination paper (Anchor Paper, St. Paul, MN, USA) was moistened using a fungicide solution Captan (2.5g/l) before four seeds per paper roll were vertically rolled. All seed rolls were placed into two liter glass beakers containing 1.4 liters of autoclaved deionized water. Growing condition settings within the growth chambers were the same as the previous GWAS study. After 14 days of growth, seedlings were removed from the growth chamber and roots were scanned for root trait measurement extraction. If not measured the same day, plants were preserved in 30% ethanol to prevent and further growth or aging. TRL was extracted from all images using the image analysis software *ARIA* (Automatic Root Image Analyzer), a high-throughput software system that can extract up to 27 root traits currently (Pace et *al.* 2014). Above ground plant material was removed from the root system prior to image capture using a high resolution scanner EPSON Expression 10000 XL.

Phenotypic data analysis

Training population phenotypic analysis has been reported in Pace *et al. (*2015). Validation experiments were carried out in a completely random design (CRD) in three experiments in the months of June and July 2014. The starting dates for experiments were June 4th, June 24th, and June 29th. All experiments were grown in the same growth chamber. All traits

data for phenotypic analysis was collected on a "plot" basis with each seed roll representing a plot. Each plot or experimental unit consisted of three seedlings being sampled, measured, and means calculated. Lsmeans of root traits analyzed were calculated based on the following linear model:

$$y_{ijk} = \mu + E_i + R_{(i)j} + G_k + e_{(i)jk}$$

Where y_{ijk} represents the observation from the ij^{th} plot, μ is the overall populations mean, E_i is the i^{th} experiment and is considered random, $R_{(i)j}$ is the j^{th} replication nested within the i^{th} experiment and is also a random effect, and G_k is the k^{th} line and is a fixed effect. All interactions with the random effect were confounded within the error $e_{(i)jk}$. The statistical software package SAS 9.3 (SAS Institute, Cary NC) was used to obtain the analysis of variance (ANOVA) table and least square means. Function PROC MIXED was implemented with type 3 sums of squares. To test whether the Long group of 100 lines and Small group of 100 lines were significantly different, a grouping term was added to the additive model as follows:

$$y_{ijkl} = \mu + E_i + R_{(i)j} + Grp_k + G_{(k)l} + e_{ijkl}$$

Where genotype is now nested within the group effect and therefore considered a random effect and *Grp* is considered a fixed effect. SAS function PROC MIXED was used to calculate expected means squared, group Lsmeans, and determine if the grouping effect was significant.

Genotypic data

Genotyping by sequencing (GBS) (Elshire *et al*. 2011) data is publicly available for the Ames Panel (Romay *et al*. 2013) and was used for genomic prediction in this study. A total of 681,257 markers distributed across the entire maize genome were available. To clean the

imputed data set, monomorphic markers, all markers with minor allele frequency < 5%, and all markers that had > 20% missing data were filtered out and were not used to train the model. A final set of 186,849 markers across the maize genome were used to train and predict performance in regards to total root length of the training and validation populations. Marker data used for GWAS were based on the entire set of 584 inbred lines and the same GBS data set and filters mentioned previously (Pace *et al*. 2015). In total, 135,311 markers distributed across all 10 chromosomes in the maize genome were used for GWAS.

Genome wide association study validation

GWAS methods used in the present study have been reported previously by Pace *et al.* (2015). The association mapping population used for validation combined both the training and validation population totaling 584 inbred lines. Population structure (Q matrix) was estimated from a reduced number of 1023 random markers across the maize genome. The software program Structure 2.3.4 (Pritchard 2000) was used with parameter settings of a burn in length of 500,000 followed by 500,000 iterations for each of the clusters (K) from 1-15. Each K was run five times. An admixture model was applied with independent allele frequencies. This model allows for the possibility that lines may have mixed ancestry in more than one sub-population (K). The most probable value for K was selected using an ad hoc method as explained in (Evanno *et al*. 2005), which is based on the ordering rate of change of P(X|K). The program called SPAGeDi (Hardy and Vekemans 2002) was used to calculate the Loiselle kinship matrix coefficients between lines (K matrix). Both Q and K matrix were fit to a mixed linear model (MLM) utilizing the program GAPIT (Genome Association and Prediction Integrated Tool- R package) (Lipka *et al.* 2012). All model parameters for GWAS are found in Pace *et al.* 2015. A General Linear Model (GLM) was also used fitting just the Q matrix to mirror the methods used

previously. Program TASSEL 4.0 was used to implement the GLM procedure. To account for multiple testing, the same stringent level as Pace *et al.* (2015) was applied in order to call significant loci $p<5.3 \times 10^{-7}$ based on an α level of 0.05.

Ridge regression BLUP

To perform genomic prediction within the current experiment, RR-BLUP (Wittaker *et al.* 2000 and Endelman 2011) was used. All genomic prediction procedures performed herein were executed using the software program R v. 3.0.2 (R Development Core Team 2013). The developed R functions are freely available online within the rrBLUP package (Endelman 2011). The mixed model (Henderson 1984) used for the RR-BLUP procedure is defined as follows:

$$Y = X\beta + Wu + e$$

Where Y represents an N x 1 vector of phenotypic Lsmeans where N represents the number of individuals in the training population, X is an identity matrix and β is the overall average of the training population. W represents an N x Nm marker matrix where Nm represents the number of markers used for GP. The u represents an Nm x 1 marker effect vector with e representing the N x 1 vector of residuals. For RR-BLUP we assume that marker effects are normally distributed $u \sim N(0,\sigma^2_u)$ with equal marker variance (σ^2_u) across the whole genome. Based on this we assume all markers have common variance (Meuwissen *et al*., 2001) and shrinkage for marker effects are equal for all markers in order to reduce estimation error. Accuracy of predictions for cross-validation and for GP between the 384 line training population and 200 line validation population is based upon Pearson correlation (r) between predicted GEBVs and observed Lsmeans ($r_{\hat{y},y}$). The adjusted prediction accuracy was estimated by taking Pearson correlation ($r_{\hat{y},y}$) between predicted GEBVs and observed Lsmeans and dividing that by the square root of

broad sense heritability (H^2) $r_{(\hat{g},g)} = \frac{r_{\hat{y},y}}{\sqrt{H^2}}$. Heritability estimates were obtained as described in Pace *et al.* 2015. The training population heritability was used as this was a randomly selected population, the validation was selected for and therefore not a good estimate of heritability for a trait. Prediction accuracies are reported as the Pearson correlation between predicted and observed root lengths to avoid added error from the heritability calculation that could inflate actual predictability. Adjusted prediction accuracies are used for comparisons to unadjusted prediction accuracies.

Acknowledgements

USDA NIFA project IOW01018 providing the experimental materials for this study. JP was supported by NSF GK-12 project and R.F. Baker Center for Plant Breeding at ISU. Advice and input for analysis of GP results ny Dr. Jack Dekkers at Iowa State University is much appreciated.

References

Aiken R.M., Smucker A.J.M. (1996) ROOT SYSTEM REGULATION OF WHOLE PLANT GROWTH. *Annual Review of Phytopathology,* **34:** 325-346.

Asoro, F.G., Newell, M.A., Beavis, W.D., Scott, M.P. and Jannink, J.-L. (2011) Accuracy and Training Population Design for Genomic Selection on Quantitative Traits in Elite North American Oats. *The Plant Genome Journal***, 4**, 132.

Bernardo, R. (2008) Molecular Markers and Selection for Complex Traits in Plants: Learning from the Last 20 Years. *Crop Sci.,* **48**, 1649-1664.

Bernardo, R. and Yu, J. (2007) Prospects for Genomewide Selection for Quantitative Traits in Maize. *Crop Sci.,* **47**, 1082-1090.

Bucksch, A., Burridge, J., York, L.M., Das, A., Nord, E., Weitz, J.S. and Lynch, J.P. (2014) Image-based high-throughput field phenotyping of crop roots. *Plant physiology*, **166**, 470-486.

Casa A.M., Pressoir G, Brown P.J., Mitchell S.E., Rooney W.L., Tuinstra M.R., Franks C.D., and Kresovich S. (2007) Community Resources and Strategies for Association Mapping in Sorghum. *Crop Sci.,* **48,** 30-40.

Clark, S.A., Hickey, J.M. and Werf, J.H.J. (2011) Different models of genetic variation and their effect on genomic selection *Genetics Selection Evolution*, **43**, 18.

Daetwyler, H.D., Calus, M.P.L., Pong-Wong, Campos, G. and Hickey, J.M. (2013) Genomic Prediction in Animals and Plants: Simulation of Data, Validation, Reporting, and Benchmarking. *Genetics,* **193,** 347-365.

Eathington, S.R., Crosbie, T.M., Edwards, M.D., Reiter, R.S. and Bull, J.K. (2007) Molecular Markers in a Commercial Breeding Program. *Crop Science*, **47**, S-154.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S. and Mitchell, S.E. (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS one*, **6**, e19379.

Endelman J.B. (2011) Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *The Plant Genome*, **4**, 250-255.

Evanno, G., Regnaut, S. and Goudet, J. (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol*, **14**, 2611-2620.

Hardy, O.J. and Vekemans, X. (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2**, 618-620.

Hayes, B.J., Bowman, P.J., Chamberlain, A.C., Verbyla, K. and Goddard, M.E. (2009) Accuracy of genomic breeding values in multi-breed dairy cattle populations. *Genet Sel Evol*, **41**, 51.

Heffner, E.L., Sorrells, M.E. and Jannink, J.-L. (2009) Genomic Selection for Crop Improvement. *Crop Science*, **49**, 1.

Heslot, N., Yang, H.-P., Sorrells, M.E. and Jannink, J.-L. (2012) Genomic Selection in Plant Breeding: A Comparison of Models. *Crop Science*, **52**, 146.

Hochholdinger, F. ed (2009) *Handbook of Maize: Its Biology*.

Hund, A., Reimer, R. and Messmer, R. (2011) A consensus map of QTLs controlling the root length of maize. *Plant and Soil*, **344,** 143-158.

Hund, A., Trachsel, S. and Stamp, P. (2009) Growth of axile and lateral roots of maize: I development of a phenotying platform. *Plant and Soil*, **325**, 335-349.

Jessica E. Rutkoski, E.L.H., Mark E. Sorrells (2010) Genomic selection for durable stem rust resistance in wheat. *BGRI 2010 Technical Workshop*.

Jonathan K. Pritchard (2000) Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 945-959.

Kumar, B., Abdel-Ghani, A.H., Pace, J., Reyes-Matamoros, J., Hochholdinger, F. and Lubberstedt, T. (2014) Association analysis of single nucleotide polymorphisms in candidate genes with root traits in maize (Zea mays L.) seedlings. *Plant science*, **224**, 9-19.

Lee M., Sharapova N., Beavis W.D., Grant D., Katt M., Blair D., and Hallauer A. (2002) Expanding the genetic map of maize with the intermated B73 x Mo17 *(IBM)* population. *Plant Molecular Biology,* **48:** 453-461.

Lipka, A.E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P.J., Gore, M.A., Buckler, E.S. and Zhang, Z. (2012) GAPIT: genome association and prediction integrated tool. *Bioinformatics*, **28**, 2397-2399.

Liu, S., Yeh, C-T., Tang, H. M., Nettleton, D. and Schnable, P. (2012) Gene mapping via Bulked Segregant RNA-Seq (BSR-Seq). *PLoS ONE,* **7,** e36406.

Liang, L., Jaiswal, P., Hebbard, Clair., Avraham, S., Buckler, E.S., Casstevens, T., Hurwitz, B., McCouch, S., Ni, J., Pujar, A., *et al.* (2007) Gramene: a growing plant comparative genomics resource. *Nucleic Acids Research,* **36,** D947-D953.

Lorenzana, R. and Bernardo, R. (2009) Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. *Theoretical and Applied Genetics*, **120**, 151-161.

Luan, T., Woolliams, J.A., Lien, S., Kent, M., Svendsen, M. and Meuwissen, T.H. (2009) The accuracy of Genomic Selection in Norwegian red cattle assessed by cross-validation. *Genetics*, **183**, 1119-1126.

Lynch J, (1995) Root Architecture and Plant Productivity. *Plant Physiology,* **109:** 7-13.

Maher B. (2008) Personal Genomes: The case of the missing heritability. *Nature,* **456:** 18-21.

Makay, I. and Powell W. (2007) Methods for linkage disequilibrium mapping in crops. *Trends in Plant Science,* **12**, 57-63.

Malamy, J.E. (2005) Intrinsic and environmental response pathways that regulate root system architecture. *Plant, Cell & Environment*, **28**, 67-77.

Meuwissen, T.H.E., Hayes, B.J. and Goddard, M.E. (2001) Prediction of Total Genetic Value Using Genome-Wide Dense Maker Maps. *Genetics* **157**, 1819-1829.

Michelmore, R.W., Paran, I. and Kesseli R.V. (1991) Identification of markers linked to disease resistance genes by bulked segregant analysis: a rapid method to detect markers in

specific genomic regions using segregating populations. *Proc Natl Acad Sci USA,* **88**, 9828-9832.

Morris G.P., Punna R., Deshpande S.P., Hash T.C., Shah T., Upadhayaya H.D., Kresovich S. et al. (2013) Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proc Natl Acad Sci USA,* **110**, 453-458.

Muñoz-Amatriaín, M., Cuesa-Marcos, A., Endelman, J.B., Comadran, J., Bonman, J.M., Bockelman, H.E., Chao, S., Russel, J., Waugh, R., Hayes, P.M., Muehlbauer, G.J. (2014) The USDA Barley Core Collection: Genetic Diversity, Population Structure, and Potential for Genome-Wide Association Studies. *PLoS ONE,* **9,** e94688.

Pace, J., Gardner, C., Romay, C., Ganapathsybrumanian, B. and Lubberstedt, T. (2015) Genome-wide association analysis of seedling root development in maize (Zea mays L .*). BMC genomics*, **16**, 47.

Pace, J., Lee, N., Naik, H.S., Ganapathysubramanian, B. and Lübberstedt, T. (2014) Analysis of Maize (*Zea mays L.*) Seedling Roots with the High-Throughput Image Analysis Tool *ARIA* (Automatic Root Image Analysis). *PloS one*, **9**, e108255.

R. Core Team. (2013) A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Australia,* http://www.R-project.org/.

Ribaut, J.-M., Fracheboud, Y., Monneveux, P., Banziger, M., Vargas, M. and Jiang, C. (2007) Quantitative trait loci for yield and correlated traits under high and low soil nitrogen conditions in tropical maize. *Molecular Breeding*, **20**, 15-29.

Romay, M.C., Millard, M.J., Glaubitz, J.C., Peiffer, J.A., Swarts, K.L., Casstevens, T.M., Elshire, R.J., Acharya, C.B., Mitchell, S.E., Flint-Garcia, S.A., McMullen, M.D., Holland, J.B., Buckler, E.S. and Gardner, C.A. (2013) Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biol*, **14**, R55.

Sallam, A.H., Endelman, J.B., Jannink, J.-L. and Smith, K.P. (2014) Assessing Genomic Selection Prediction Accuracy in a Dynamic Barley Breeding Population. *The Plant Genome*.

Sekhon, R.S., Lin, H., Childs, K.L., Hansey C.N., Buell C.R., De Leon, N., Kaeppler, S.M. (2011) Genome-wide atlas of transcription during maize development. *Plan J.* **66**(4): 553-563.

Spindel, J., Begum, H., Akdemir, D., Virk, P., Collard, B., Redona, E., Atlin, G., Jannink, J., McCouch, S.R. (2015) Genomic Selection and Association Mapping in Rice (*Oryza sativa*): Effect of Trait Genetic Architecture, Training Population Composition, Marker Number and Statistical Model on Accuracy of Rice Genomic Selection in Elite, Tropical Rice Breeding Lines. *PLoS Genet,* **11**(2): e1004982.

Stupar R.M., Specht J.E. (2013) Insights from the soybean (*Glycine max* and *Glycine soja*) genome: past, present, and future. *Adv. Agron.* **118:** 117-204.

Taylor, J.R. (1997) An Introduction To Error Analysis The Study Of Uncertainties In Physical Measurements: Second Edition. *University Science Books*, 128-129.

Tuberosa, R. (2012) Phenotyping for drought tolerance of crops in the genomics era. *Frontiers in Physiology*, **3**.

Venuprasad, R., Dalid, C. O., Valle, M. D., Zhao, D., Espiritu, M., Sta Cruz, M. T., Ameante, M., Kumar, A. and Atlin, G. N. (2009) Identification and characterization of large-

effect quantitative trait loci for grain yield under lowland drought stress in rice using bulk-segregant analysis. *Theor Appl Genet,* **120,** 177-190.

Whittaker, J.C., Thompson, R. and Denham, M.C. (2000) Marker-assisted selection using ridge regression. *Genet. Res., Camb*., 249-252.

Yu, J., Pressoir, G., Briggs, W.H., Vroh Bi, I., Yamasaki, M., Doebley, J.F., McMullen, M.D., Gaut, B.S., Nielsen, D.M., Holland, J.B., Kresovich, S. and Buckler, E.S. (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet***, 38**, 203-208.

Zeng, J., Pszczola, M., Wolc, A., Strabel, T., Fernando, R.L., Garrick, D.J. and Dekkers, J.C. (2012) Genomic breeding value prediction and QTL mapping of QTLMAS2011 data using Bayesian and GBLUP methods. *BMC proceedings*, **6** Suppl 2, S7.

Zhong, S., Dekkers, J.C., Fernando, R.L. and Jannink, J.L. (2009) Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a Barley case study. *Genetics*, **182**, 355-364.

Table 1. Predicted vs observed TRL means of selected extreme genotypes

Group	Source of data	Lsmeans	St. Error
Short Group	Predicted	100.92 cm	-
Short Group	Observed	176.44 cm	9.14
Long group	Predicted	246.06 cm	-
Long group	Observed	295.11 cm	9.10

Table 2. Average ranking accuracy and adjusted ranking accuracy for three root traits TRL, SEL, PRL

Population size	Trait	Avg ranking Accuracy $\rho_{\hat{y},y}$	Avg Adjusted Accuracy $\rho_{\hat{g},g}$
200	TRL	r=0.55	r=0.84
200	SEL	r=0.54	r=0.83
200	PRL	r=0.38	r=0.72

TRL= Total Root Length, SEL=Secondary Root Length, PRL=Primary Root Length

Table 3. Average prediction accuracy and adjusted accuracy for three root traits TRL, SEL, PRL

Population size	Training population size	Trait	Avg Phenotypic Accuracy $r_{\hat{y},y}$	Avg Adjusted Accuracy $r_{\hat{g},g}$	St. Deviation
384	234	TRL	r=0.42	r=0.65	0.056
384	234	SEL	r=0.43	r=0.67	0.055
384	234	PRL	r=0.44	r=0.84	0.051
584	350	TRL	r=0.54	r=0.84	0.035
584	350	SEL	r=0.32	r=0.49	0.049
584	350	PRL	r=0.31	r=0.58	0.047

TRL= Total Root Length, SEL=Secondary Root Length, PRL=Primary Root Length

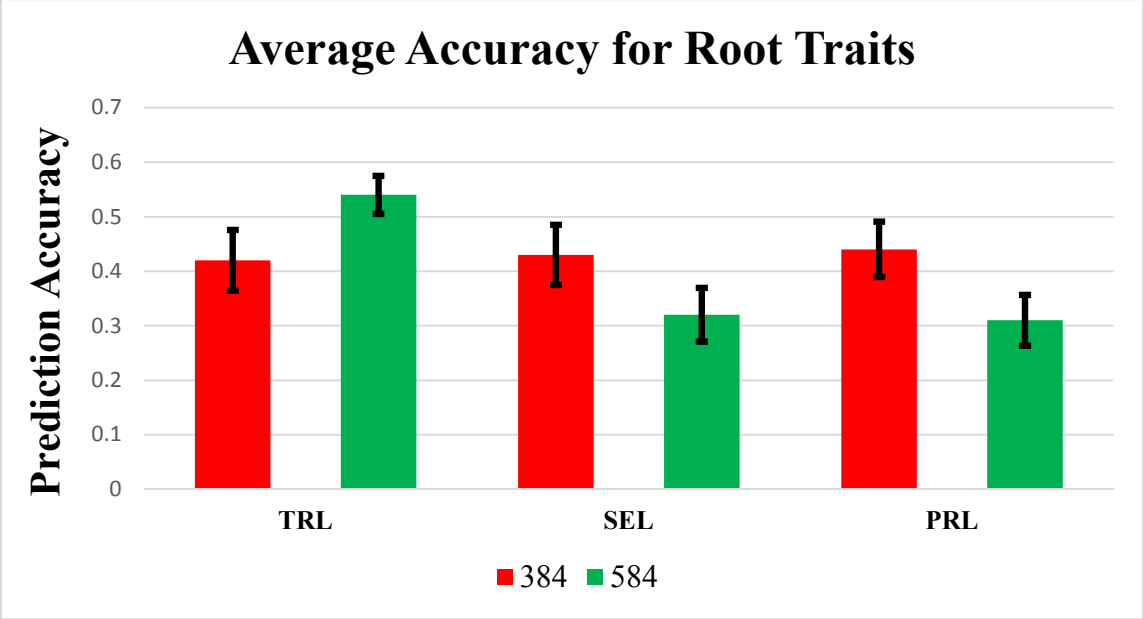


Figure 1. Average 60/40 cross-validation accuracy for 384 and 584 line populations, error bars represent st. dev of accuracy estimates for all 500 iterations of RR-BLUP

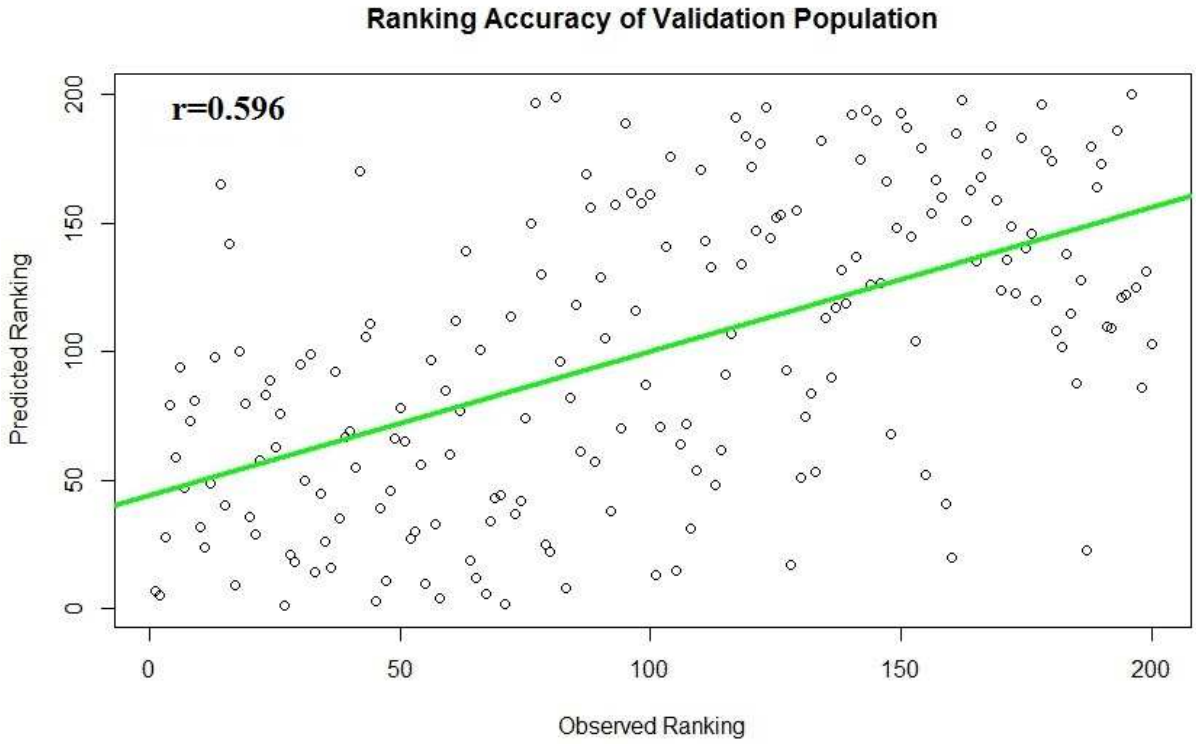


Figure 2. Ranking accuracy for TRL for 200 line validation population with ranks ranging from 1-200, one being the longest roots and 200 being the shortest roots.

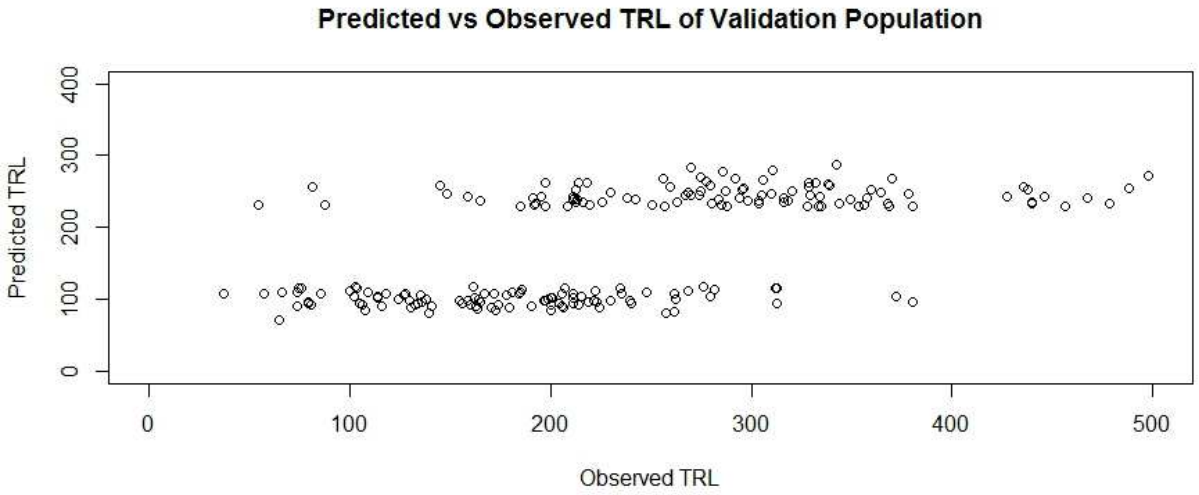


Figure 3. Validation of predictions made based on TRL within the validation population. Overall prediction accuracy was estimated at r=0.594.

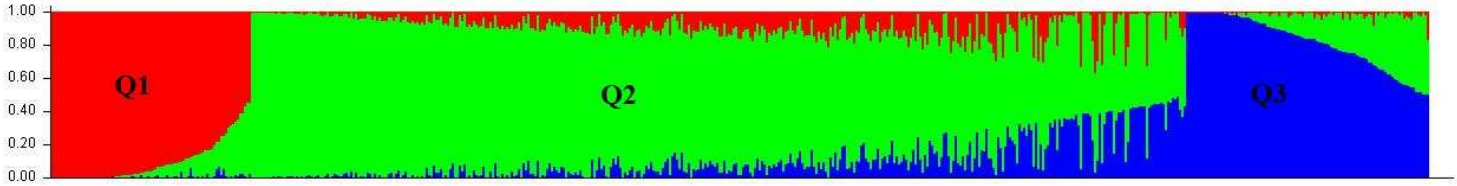


Figure 4. Population stratification identified using STRUCTURE 2.3.4 the 584 line population after adding in the validation population. Q1, Q2, and Q3 represent their respective subpopulations.

CHAPTER SIX

GENERAL CONCLUSIONS AND FUTURE PERSPECTIVES

The goal of this project was to explore root trait variation and to better understand the genetic architecture controlling root development within maize. This was accomplished by analyzing seedling root architecture traits within two separate association analysis inbred line panels, and using the Ames panel population to complete a genomic prediction study. Loci found significantly associated with seedling root traits were identified at the gene level within a candidate gene association study within a 74 line association study (AS) panel as well as a GWAS within a larger 384 line association mapping population. These loci represent putative candidates for future study and validation. A new expandable tool called *ARIA* allowing high-throughput root trait imaging was developed to facilitate studies herein as well as future studies done to better understand root architecture. Genomic prediction (GP) was tested for root architecture traits in order to determine whether extreme genotypes for a moderately heritable trait can be predicted with sufficient accuracy. Within the candidate gene association mapping study, candidate genes *Rtcl, Rth3, Rum1,* and *Rul* were re-sequenced within 74 diverse inbred lines. Root traits were extracted from 6, 10, and 14 day old seedlings and marker-trait associations were calculated. A total of 51 SNP trait associations were identified. These significant SNPs within root development genes form the basis for putative functional markers for breeding purposes. For GWAS, a subset of 384 inbred lines of the Ames panel was grown for 14 days, with root traits extracted using the software program *ARIA*. Marker trait- associations were calculated using both MLM and GLM, identifying 4 and 287 marker trait associations, respectively, for each model used. Within the loci identified by GLM, 17 were associated with multiple root traits. Putatively associated SNPs located within or near gene models with regard to

the B73 reference genome represent the best candidates for genes having an effect on root development in maize. Gene model GRMZM2G153722 on chromosome 4 contained 12 significant markers for root trait DIA. This model represents the best candidate for a root gene as it has homology for a root tip development gene in Arabidopsis.

The GWAS panel was used as a training population for GP based on TRL. A RR-BLUP prediction model was implemented to calculate marker effects and predicted BLUPs based on genotypic information in the validation population. A validation population of 200 inbred lines from the Ames diversity collection was selected based on predicted extreme genotypes with regard to the longest and shortest predicted root lengths. This study showed that GP was able to differentiate two groups with statistically different TRL with even larger observed than predicted difference for the TRL means. Genotype ranks were also predicted with moderate accuracy with Spearman correlation of ρ=0.55.

Because the studies described herein were performed at the seedling stage and within a controlled environment, they are only a first step towards a better understanding of the genetic control and predictability of root traits in maize. Future studies need to address how environments affect GWAS as well as GP of maize root traits. For practical purposes, it has been noted that breeders do not currently use root architectural traits directly as a selection criterion. With the use of GS and GP technologies, a reduction in resources required to collect phenotypic data may make use of root traits for selection more realistic. It also needs to be better understood, how environmental variation affects prediction accuracies more generally. Moreover, what is the contribution of roots to grain yield? Can specific root architectural traits be correlated to yield, and will root traits be more important in nutrient deficient environments? There are still many questions to be answered before roots are a common selectable component

to breeding programs in maize and other crop species. Studies answering all of these questions are ongoing but preliminary results do show promise with newer technologies in genotyping and phenotyping becoming increasingly readily available.

ACKNOWLEDGEMENTS

I would like to take this chance to thank those who have helped me not only write this thesis, but have helped me throughout my graduate career with direction, teaching and support in all facets. First I thank my advisor Thomas Lübberstedt and the entire Lübberstedt lab for their efforts in helping me with advice and support throughout my research. Next, the remaining members of my POS committee, all of whom have always been willing to give helpful advice and left an open door I have felt welcome in. I would also like to thank my collaborators on my projects, Baskar Ganapathysubramanian, Cinta Romay, and Xiaoqing Li. Finally, I would like to thank my family for their love and support. These thanks go especially to my wife Jesa, who has stuck by me and encouraged me with love and patience through all trials and tribulations I have faced throughout my time at Iowa State University. To all who have helped me throughout this journey, thank you.