2017

Genome wide association study of seed and seedling root traits in sunflower

Ze Tian Fang
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/etd

Part of the Agricultural Science Commons, Agriculture Commons, and the Agronomy and Crop Sciences Commons

Recommended Citation

Fang, Ze Tian, "Genome wide association study of seed and seedling root traits in sunflower" (2017). Graduate Theses and Dissertations. 16125.
https://lib.dr.iastate.edu/etd/16125

This Thesis is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
Genome wide association study of seed and seedling root traits in sunflower

by

Ze Tian Fang

A thesis submitted to the graduate faculty in partial fulfillment of the requirements for the degree of MASTER OF SCIENCE

Major: Plant Breeding

Program of Study Committee:
Jessica Barb, Co-Major Professor
Thomas Lübberstedt, Co-Major Professor
Allen Knapp

The student author, whose presentation of the scholarship herein was approved by the program of study committee, is solely responsible for the content of this thesis. The Graduate College will ensure this thesis is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University
Ames, Iowa
2017

Copyright © Ze Tian Fang, 2017. All rights reserved.
# TABLE OF CONTENTS

## CHAPTER 1. LITERATURE REVIEW

- Root Importance ................................................................................................................. 1
- Overview ............................................................................................................................... 1
- Root variability ..................................................................................................................... 1
- Sunflower Root System ........................................................................................................ 2
- Phenotyping of Roots .......................................................................................................... 3
- Sunflower ............................................................................................................................ 7
- Introduction ......................................................................................................................... 7
- Genetics ............................................................................................................................... 8
- QTL & Association Mapping ............................................................................................... 11

## CHAPTER 2. GENOME WIDE ASSOCIATION STUDY OF SEED AND SEEDLING ROOT TRAITS IN SUNFLOWER

- Abstract .............................................................................................................................. 13
- Introduction ......................................................................................................................... 13
- Material and Methods ........................................................................................................ 15
- Results ................................................................................................................................ 18
  - Seed and seedling root phenotypes .................................................................................. 18
  - Pearson correlations ........................................................................................................ 19
  - PCA analysis .................................................................................................................... 19
  - Marker and GWA ............................................................................................................. 20
- Discussion ............................................................................................................................ 21
- Conclusion ........................................................................................................................... 24
- Tables and figures ................................................................................................................. 25
- Supplemental Figures .......................................................................................................... 32
- BIBLIOGRAPHY ............................................................................................................... 36
CHAPTER 1. LITERATURE REVIEW

Root Importance

Overview

Roots are responsible for anchorage, nutrient and water absorption, and hormone production. Primary roots contain a meristem, similar to a shoot apical meristem. Roots demonstrate geotropism, negative phototropism (i.e., the ability to grow away from light), and hydrotropism (i.e., the ability to grow towards water). Monocot root systems are fibrous while dicots root systems usually have a primary root that is gravitropic with lateral secondary roots that can be anti-gravitropic (Torrey 1976, Ge, Rubio and Lynch 2000, Grossnickle 2005).

Root variability

Variation in root architecture could be used by breeders to improve root traits (e.g., better nutrient and water absorption) in crop plants. In maize (Zea mays) root density affects nitrogen uptake (Beebe et al. 2006). In barley (Hordeum vulgare), wild accessions have fewer seminal roots with stronger geotropism in comparison to modern cultivars that have more seminal roots and thus a larger root system that can exploit more soil volume (de Dorlodot et al. 2007).

In sunflower (Helianthus annuus L.), root growth was characterized by comparing growth rates in the primary root vs. lateral roots in six wild species intercross lines and two cultivated hybrids (i.e., one semi-dwarf and one with standard height) (Seiler 1994). Growth rates in one intercross line exceeding that observed in the cultivated hybrids indicating that
wild species may contain genetic diversity that could improve root growth in cultivated hybrids. Root architecture traits were highly correlated (e.g., \( r \approx 0.85 \) for primary root length and lateral root number) suggesting that these traits may be controlled via a common genetic pathway, which could simplify future breeding efforts. Roots of the semi-dwarf hybrid were the longest at the seedling stage, but were no longer the longest at maturity demonstrating the need for future studies to connect early root system growth to later root system growth.

Sunflower Root System

In 1926, root growth of a Russian sunflower cultivar was characterized under upland and lowland field conditions under different planting densities. Plants in both lowland and upland conditions produced taproots that reached depths of 38.1 to 45.7 cm in the first five weeks with lateral roots extending 30.5 cm outward from the base of the plant. At 10 weeks, plants were ~213.0 cm tall and had taproots that were 167.7 cm long. Most of the lateral roots were located in the top 45.7 cm of the soil, but some lateral roots were found as deep as 121.9 cm. Plant height, leaf number, and root size were reduced in plants grown at a higher density (i.e., 5.0 cm between plants) relative to plants grown with 20.3 or 81.4 cm between plants. Despite this reduction in overall growth, roots reached a depth of 182.9 cm (Weaver and Bruner 1926, Sadras et al. 1989).

In 1986, Jaafar et al. examined root depth in field grown sunflowers using the core-break method. Peak root growth occurred during early plant development (i.e., V2 to V10) with root growth ceasing after R8 (i.e., seed maturity). The deepest root was located at 2.7 m, which was similar to the report of ~274.0 cm in Weaver et al. in 1926. This exceeds that
observed in corn (~180 cm), soybean (~152 cm), and sorghum (~180 cm) (Weaver and Bruner 1926, Hoogenboom, Huck and Peterson 1987). Although similar studies have not been conducted using modern hybrids, it is likely that this level of variation exists and could be used by breeders to create hybrids with better chemotaxis, better soil exploration, and higher growth rates, which could improve overall nutrient and water absorption.

Phenotyping of Roots

Despite the importance of roots to the overall productivity of plants, selection programs have not been conducted to specifically improve roots traits. This is likely attributed to the difficulties associated with accurately phenotyping roots in a timely, cost efficient, non-destructive, and repeatable manner. Phenotyping roots in the field is labor intensive and time-consuming. Excavation procedures often leave lateral roots behind even with the best precautions, which inevitably decreases the reliability and accuracy of the results. Furthermore, excavation methods are destructive so they are not repeatable, which increases cost and space requirements. Non-invasive methods such as ground radar have been used to characterize tree roots with some success, but the low resolution of the images did not provide adequate information for selection (Butnor et al. 2003).

The core-break method is a quick method for examining primary root length in plants (Bennie, Taylor and Georgen 1987). This method was discussed by Bohm in 1979 and used to study root development in field grown sunflower in 1985. In Bennie et al. 1987, the core-break method was compared with a procedure where roots were excavated, washed and measured. The core-break method used a correlation equation for estimating root density (i.e., $L = bN$, $L$ = root length per volume, $b$ = coefficient, $N$ = number of roots per intersection
Four types of plants were evaluated using both methods to determine the appropriate coefficient \( b \) for different crops. The results showed that \( b = 2 \) was appropriate for estimating root density in dicots, while \( b = 1.5 \) was more suitable for monocots. The correlation between the methods was different depending on the species tested (e.g., sunflower \( R^2 = 0.54 \), kochia \( R^2 = 0.74 \)), due to differences in lateral root growth. This study also noted that human bias and errors affected the results of the core break method with the biggest limitation stemming from the high frequency of broken soil cores. Despite these problems, the authors concluded that the core-break method was a valid method for the rapid estimation of root density once the coefficient \( b \) for a specific crop was validated.

Non-invasive methods provide a better view of root growth in real-time. These methods utilize different growing media such as greenhouse soil mix, sand, clear gel, and water. Each medium has distinct advantages for viewing and measuring plant roots, with most demonstrating some ability to predict root growth in a field setting (Mian et al. 1993, Price, Tomos and Virk 1997, Price and Tomos 1997, Toda, Koyama and Hara 1999). In Price et al. root length and thickness were measured over time in 300 hydroponically grown rice (\textit{Oryza sativa}) plants. These results were highly correlated \( r > 0.80 \) with a field study on root growth of 12 \textit{Oryza} species conducted by Loresto et al. 1983. Despite the high correlation between these two studies, it is unknown whether data from hydroponic experiments will correlate as well in other crops and for other root architecture traits (e.g., primary root length, root column diameter, and root volume). Furthermore, data from hydroponic systems tend to be poor predictors of plant behavior under drought stress (Mian 1993).
Today, 2D imaging programs (e.g., WinRhizo® and ARIA®) are commonly used for phenotyping roots. These systems are easy to use and produce consistent results (Pace et al. 2014, Pace et al. 2015), which can be used for genome wide association (GWA) studies. In 2012, Kumar et al. examined root morphology in 10-day-old maize seedlings (i.e., 44 expired Plant Variety Protection lines and 33 public inbred lines). Significant differences among genotypes for all traits (e.g., primary root length, lateral root length, and total root length) and at all dates were observed. CV values for primary root length and crown root length decreased over time while variation in seminal root length was highly variable (i.e., CV values of 40-50) across all time points. Relative growth rates were faster from day 0 to day 10 and then decreased after the endosperm was likely depleted. Seed size and root growth were not correlated in this study. Cluster analysis identified two lines with stunted root growth. Because these lines were unrelated, it was concluded that breeders probably did not actively select for root traits in these lines and that the variation observed was randomly distributed throughout the collection. In 2014, candidate genes were sequenced in 74 lines used in the 2012 study and roots were measured using WinRhizo-Pro® 2009 (Kumar et al. 2014). Several QTL associated with early root growth were identified. Population structure based on root measurements vs. pedigree data showed that polymorphisms in root related genes were highly diverse and randomly distributed across multiple unrelated lines. In Pace et al. 2014, root growth was characterized in 384 maize inbred lines using the protocol in Kumar et al. 2012 and 2014, where four seeds of each genotype were placed in paper rolls and incubated in a growth chamber for 14 days. Root traits were captured using WinRhizo-Pro® 2009 and analyzed using ARIA® version 1.0. Root
traits were significantly different among lines. Heritability estimates ranged from 0.12 to 0.49, which was lower than the values observed by Kumar et al. 2014 (i.e., $H^2 = 0.66 - 0.90$). Population clustering based on these traits produced two large populations separating B73 and Mo17. GWA analysis using a MLM model identified four markers associated with root length and root bushiness. An additional 264 markers were identified using a GLM model. In total, these studies demonstrated that 2D images of seedling roots produced using the paper roll method were adequate for capturing genetic variation in a diverse collection of maize lines and that this variation was suitable for genetic analysis and QTL detection.

In 2011, Famoso et al. examined root architecture in rice using a 3D platform that utilized different growing media (i.e., gellan gem, sand, and hydroponics). Cameras were used to capture real time root growth of seedlings grown in containers mounted to a 360-degree rotating stand. The correlation between 3D data and 2D data across all traits was high demonstrating that their 3D platform was suitable for measuring root system architecture in rice. Other 3D imaging techniques have been used to measure root growth in other species (Trachsel et al. 2011) including computed tomography (CT) scanning (Lontoc-Roy et al. 2006) and magnetic resonance imaging (MRI) (Jahnke et al. 2009), however these methods are not widely used due to their high costs and need for specialized equipment.

Both 2D and 3D imaging platforms are useful for capturing variation in root morphology (as reviewed in Böhm 2012). 2D platforms are useful for imaging a large number of samples in a short time. They are accurate at measuring total root length (Clark et al. 2011), but they don’t account for overlapping roots as well as a 3D scan can. 3D platforms tend to mimic actual growth conditions better and can accurately characterize
overall root structure, but they are less high throughput than 2D platforms. Computed tomography scanners and laser scanners require specialized machinery and do not yet produce the resolution that is achievable with 2D scanners (Zhu et al. 2011). Roots produced under hydroponic conditions can be measured using 3D platforms, but the data may not correlate well with that observed when roots are grown in soil or in paper rolls (i.e., for prediction accuracy). In general, due to the high costs and time required for data collection, 3D platforms are better suited for smaller studies that are conducted to characterize root architecture in only a few genotypes. Finally, regardless of which platform (e.g., 2D or 3D) is used, field studies should be conducted in parallel to ascertain how root growth changes under field conditions and how the environment affects root development and architecture as this will likely impact the final realized phenotype that breeders are selecting for.

Sunflower

Introduction

Sunflower was domesticated in the United States (i.e., likely as a single event, Blackman et al. 2011) ~4800 years ago (Heiser Jr 1954, Heiser 1955, Heiser et al. 1969, Heiser 1978, Harter et al. 2005). Sunflower was then introduced into Europe by Spanish explorers during the 1500’s as an ornamental before it became an important oil seed crop in Russia in the 1930’s (Heiser 1951). Today, sunflower is an important oil seed crop both globally (i.e., 13% of total oil crop market) and in the US (i.e., market value of $555 million dollars in 2015) (USDA 2016). Sunflower oil is low in unsaturated fats (i.e., high in linoleic acid and oleic
acids). Sunflower seeds contain ~25% protein, which is suitable for animal feed and human consumption (Putt 1997, Kole 2007).

Genetics

Cultivated sunflower is diploid ($n = 17$, $2n = 34$) and has a physical map length of ~2800-3600 Mbp (Kane et al. 2011, Renaut 2017). QTL have been identified in sunflower for yield, flowering, plant height, oil content, and other quantitative traits (Hervé et al. 2001, Al-Chaarani et al. 2002, Bert et al. 2003, Ebrahimi et al. 2008, Poormohammad Kiani et al. 2009). Simple sequence repeats (SSRs) and single nucleotide polymorphisms (SNPs) are commonly used for genetic studies in sunflower (Liu et al. 2006, Mandel et al. 2011, Nambeesan et al. 2015). RNA and cDNA data have also been used for identifying interesting genes or genetic regions (Kiani et al. 2007).

In 2006, Liu and Burke used sequencing data from 16 wild sunflowers and 16 cultivated sunflowers to show that wild sunflowers had more SNPs, more haplotype blocks, and a higher rate of LD decay. This was not unexpected because domestication involved inbreeding (i.e., less SNPs) vs. out-crossing (i.e., more SNPs), which is common in wild sunflowers since they are largely self-incompatible (Rieseberg, Baird and Desrochers 1998). Based on these findings the authors suggested that association mapping approaches in sunflower should result in higher resolution and a better connection between phenotype and genotype than traditional QTL mapping approaches.

Although modern-day sunflower hybrids have retained 65-80% of the diversity in wild Helianthus annuus (Mandel et al. 2011), cultivated sunflower is less diverse than maize (Liu et al. 2006). The lack of diversity in cultivated sunflower is likely due to a strong selection
bottleneck that occurred during domestication (Burke et al. 2002, Liu et al. 2006). However, because cultivated sunflower is inter-fertile with most wild *Helianthus* species (Seiler 1992, Arias and Rieseberg 1994, Jan et al. 2008), these species could be used to supply additional variation if needed for disease resistance or yield improvement (Mandel et al. 2011).

In 2011, Mandel et al. used 34 EST-SSRs (i.e., one marker per chromosome arm) to examine population structure and genetic diversity in a panel that included cultivated inbred lines and wild sunflower accessions. SSRs were selected because they had a presumptive neutral allele frequency (i.e., the region was not under selection pressure) and mapped to regions where QTL had not been identified previously. Wild accessions were found to contain more alleles (i.e., 328 versus 230) than cultivated lines supporting previous findings (Burke et al. 2002, Liu et al. 2006) that a strong selection bottleneck occurred in sunflower during domestication and improvement. Two distinct clusters (i.e., one with wild accessions and one with cultivated lines) were identified. Wild accessions were further divided into four sub-clusters and cultivated lines were divided into two sub-clusters (i.e., restorer lines vs. the rest). Wild sunflower sub-clusters 3 and 4 contained accessions that were collected in the central US and northern Mexico. These sub-clusters were more similar to the cultivated lines in the panel, which supported previous studies (Burke et al. 2002) that concluded that sunflower was likely domesticated in the east central US. Allele diversity in the cultivated lines was 67% of the wild accessions, which was similar to the findings in Burke et al. 2002, which was based on 122 SSRs. Computational analyses were conducted to identify a core set of 288 lines which contained 87.4% of the total allele differences. Lines were selected based on the frequency of the alleles versus their rarity in
the panel. Although the selection process to create the core panel did not include phenotypic data it was based on a reasonable assumption that lines that were the most diverse for the 34 markers tested would also be diverse elsewhere in the genome.

In 2011 a high-density SNP chip was created using sequencing data from maintainer and restorer lines and cDNA extracted from roots, seeds, leaves, and disc florets (Bachlava et al. 2012). 85,063 SNPs were identified with two or more reads per genotype, with at least one polymorphic site, and an allele frequency of 0.90. SNPs in introns were discarded and final selections were made based on the GoldenGate probe design and bead type requirements, predicted gene function, and open reading frame length. The final 10,640 unigenes were selected based on uniqueness and where they mapped to on the reference genome. This assay was tested on a diversity panel of ~90 lines. Population structure analysis produced two distinct groups of restorer lines vs. everything else. The second cluster was further delineated into three groups of oil restorer lines, oil maintainer lines, and OPV/landrace lines.

The SNP chip developed by Bachlava et al. was used to conduct association mapping studies on branching and flowering time in the core panel of 288 lines described above (Mandel et al. 2013, Nambeesan et al. 2015). Phenotypic data were collected from field sites in Iowa, Georgia and British Columbia. The number of polymorphic SNPs (i.e., 5,788) was lower than the initial report of 7,640 indicating that the diversity in the diversity panel of ~90 lines (Bachlava et al. 2012) was likely greater than that that found in the elite lines used to create the chip. The moving window analysis of LD across the genome showed different rates of decay among and within linkage groups. The authors hypothesized that
regions with lower LD decay existed due to a reduction in gene recombination in these regions especially when selection favored a specific combination of genes within that region and if two or more linked genes were present. This was supported by the observation that regions with higher LD often contained QTL for disease resistance, yield, oil content, etc., which are under intense selection by breeders. Significant peaks for branching and days to flower were found on almost every linkage group (LG). Novel QTL for both traits were identified likely due to the higher resolution achieved for association mapping vs. traditional QTL mapping.

In general, the genetic potential of sunflower is under realized relative to other crops. New discoveries and additional advancements are likely achievable due to the relatively short breeding history of sunflower. The large collection of wild accessions of sunflower is a tremendous asset for researchers and breeders (Tang and Knapp 2003), and the below average LD decay rate makes association mapping a great tool for future QTL discovery.

QTL & Association Mapping

QTL and association mapping are based on linkage disequilibrium (LD). LD occurs when alleles do not segregate independently. Genes that are located on the same chromosome arm that are in close proximity to each other do not segregate independently (i.e., the likelihood that two genes are inherited together is greater than 50%). The physical closeness of genes to each other and/or their proximity to the centromere increases or decreases associations between genes over time. QTL mapping utilizes the physical linkage (where association mapping utilizes all LD events) between markers to identify associations with an observed phenotype. Mapping QTL regardless of method can help researchers
understand relationships between genotypes and phenotypes, and eventually assist with determining causal genes.

Bi-parental populations are typically used for mapping QTL. Phenotypic and genotypic data is collected and analyzed using various methods (e.g. SIM, CIM, etc.). Required marker density is influenced by the rate of LD decay (i.e., the rate at which alleles reach genetic equilibrium). The power to detect QTL using this method is moderately high, and it is capable of detecting rare alleles. QTL found using this method are sometime unique to a population and/or an environment so they may have limited usefulness in other populations (Crosses 2001).

Association mapping uses a diverse population of unrelated lines. Population structure is greater than that found for a bi-parental QTL mapping population because certain alleles are more common due to regional adaption and natural selection ( Tanksley 1993, Pritchard et al. 2000). Because LD is eroded in these lines over hundreds or thousands of years, the physical linkage size is much smaller resulting in greater resolution of a QTL’s location if marker coverage is adequate. Association mapping doesn’t capture rare alleles or novel alleles within a sub-population, because the frequency of these alleles is not amplified like it is in a bi-parental population (and are often discarded due to sub-population specificity).

QTL mapping and association mapping are both reliable methods for detecting QTL, so the usage of one vs. the other is typically dictated by the user’s objective. QTL mapping and association mapping are both dependent on a close association between a marker and the trait of interest. If this association is not based on actual physical proximity then the effect of a marker can be misleading and lead to false results (Xu 2003).
CHAPTER 2. GENOME WIDE ASSOCIATION STUDY OF SEED AND SEEDLING ROOT TRAITS IN SUNFLOWER

Abstract

Cultivated sunflower (*Helianthus annuus* L.) is an internationally important crop harvested for seed oil and confectionary purposes. Sunflower unlike many other crops, has an extensive root system that can provide drought avoidance under water limiting conditions in the field. This genome wide association (GWA) study utilized 2D images of seeds and seedling roots from a diverse panel of sunflower genotypes (n = 288 lines) for QTL detection. The subsequent analyses of these images revealed vast phenotypic variation in total root length (mean = 169 cm, SD = 81 cm) and primary root length (mean = 29 cm, SD = 7 cm) in 10-day-old seedlings. ANOVA and PCA based on population information assigned by STRUCTURE indicated significant differences between restorer lines and maintainer lines for both seedling root and seed traits. In total, 29 unique markers associated with seed size and shape and seedling root traits were identified. This study provides the groundwork for identifying markers that breeders can use for marker assisted selection to improve sunflower seedling emergence and establishment.

Introduction

Cultivated sunflower (*Helianthus annuus* L.) is an internationally important crop harvested for seed oil, confectionary uses, and as an animal feed. In 2015, 0.73 million hectares of sunflower were harvested in the US with an estimated value of $555 million (USDA 2016). Rain-fed oilseed sunflowers make up 80% of the acreage planted in the Dakotas, Minnesota, Kansas, and Texas (USDA 2016). Sunflower plants produce extensive
root systems with primary roots capable of reaching a depth of 2 m and lateral roots extending up to 1.5 m (Weaver and Bruner 1926, Sadras et al. 1989). This extensive root system potentially enables sunflower plants to avoid water stress by gaining access to deeper soil profiles, which may contain more water. Root architecture influences nutrient uptake, water availability, and plant anchorage, which are all crucial components in determining plant productivity (Gardner 1964, Barley 1970, Sanguineti et al. 1998, de Dorlodot et al. 2007, Comas et al. 2013). Despite the importance of root architecture, it is less understood relative to above ground plant architecture due to the difficulties associated with phenotyping.

Quantitative trait loci (QTL) and association mapping approaches are useful for investigating complex traits like root architecture (Tuberosa et al. 2003, Song et al. 2007, Courtois et al. 2013), which is typically quantitatively controlled and influenced by the environment (McMichael and Quisenberry 1993, Price et al. 1997, Loudet et al. 2005). Genome wide association (GWA) studies utilize linkage disequilibrium (LD) to detect non-random association between genetic markers and phenotypes (Risch and Merikangas 1996). Unlike traditional linkage mapping, which often uses a bi-parental population, GWA studies use a diverse panel of individuals with varying levels of genetic relatedness. The LD in these collections has undergone more cycles of recombination thus allowing for greater resolution of QTL. Spurious results can occur when subpopulations share common alleles due to natural mutation and local adaptation (Korte and Farlow 2013), thus the inclusion of a kinship matrix and population structure (Pritchard et al. 2000, Yu et al. 2006) is recommended even though this may result in the omission of population specific alleles or
rare alleles that are confounded by population structure. GWA studies require adequate marker coverage and ample field space for evaluating a diverse panel of genotypes.

GWA studies are useful for identifying QTL in a wide variety of systems (i.e., from detecting human health risks to flowering time in plants) (Hirschhorn et al. 2002, Zhu et al. 2008). Plant breeders often utilize markers closely linked to GWA-detected QTL for marker assisted selection (MAS) to track and select for desired traits (Batley and Edwards 2007, Xu and Crouch 2008). In sunflower, QTL for resistance to Sclerotinia head rot, flowering time, axillary branching, oil content, seed size, and plant height (Tang et al. 2006, Fusari et al. 2012, Cadic et al. 2013, Mandel et al. 2013, Nambeesan et al. 2015) have been identified, but QTL for root traits are lacking. In the US, a majority of oilseed sunflower is planted in the Dakotas in areas with less fertile soils and limited rainfall (Supplemental Figure 1), therefore QTL for root traits could be useful for producing more resilient hybrids that yield well in these less favorable environments. The objective of this study was to identify QTL for seed and seedling root traits in a panel of diverse sunflower genotypes grown in a controlled environment. These results will establish a foundation for future experiments that will provide data to plant breeders for use in MAS programs.

Material and Methods

Plant materials: Four seed lots (i.e., 2013 rep 1 and rep 2 and 2014 rep 1 and rep 2; each harvested from individual plants = biological replicates) of a sunflower association mapping population (SAM, n = 288 lines) were produced at the North Central Regional Plant Introduction Station (NCRPIS) in Ames, Iowa in 2013 and 2014. The SAM population includes oil and non-oil maintainer (i.e., HA lines) and restorer (i.e., RHA lines) lines, wild-
introgressed lines, and open pollinated varieties (OPVs) derived from breeding collections maintained by the USDA and the French National Institute for Agricultural Research (INRA) (Mandel et al. 2011). This population represents 90% of the allelic diversity from an original collection of 433 genotypes developed in 2011.

Seed phenotyping: Each seed lot was weighed (i.e., 100-seed weight) and imaged using a flatbed scanner (Epson® Expression 10000XL). Seed images were analyzed using Smartgrain® (Tanabata et al. 2012) (Table 1).

Seed germination: Seeds were placed in clear plastic boxes (24.7 cm W X 16.5 cm L X 4.4 cm H) containing two pieces of steel blue germination blotter paper (Anchor Paper Co.) and germinated in a growth chamber (23/18°C, 16/8h light/dark cycle) (Gay, Corbineau and Côme 1991). A completely randomized design was used with each genotype consisting of ten seeds/experimental unit. The experiment was replicated three times for 2013 seed lots (i.e., 2013rep1 and 2013rep2) and twice for 2014 seed lots (i.e., 2014rep1 and 2014rep2). Germinated seeds (i.e., 2 mm protrusion of the root) were counted and then discarded. Observations began at 24 hours and were recorded every eight hours for a total of five days. Average germination time (i.e., the sum of germination time of all germinated seeds divided by the number of seeds germinated) (Orchard 1977) was calculated for each genotype and used to synchronize subsequent root phenotyping experiments.

Root phenotyping: Each seed lot was evaluated using an augmented design (i.e., independent of growth chamber) and replicated once. Six boxes were placed in a growth chamber with three boxes per shelf. Each box contained 51 genotypes and three check varieties (i.e., RHA363, RHA374, and HA89 all produced in 2013). Checks were selected
based on their genotype (HA vs. RHA), root architecture (longer total root length ~160 cm vs. shorter total root length ~130 cm), and germination time (30 hours, 43 hours, and 58 hours, respectively). Five uniform seeds of each genotype were placed (i.e., 0.8 - 1 cm from the top edge) between two pieces of germination paper (34.9 cm W X 30.5 cm L, regular weight seed germination paper, Anchor Paper Co.) that were supported by two pieces of steel blue germination blotter. Each box was divided into nine sections to ensure that each set of blotter papers was adequately supported. Each box was filled with DI water and covered with Glad® Cling Wrap to maintain humidity levels above 95%. On day three, the covers were removed to prevent contact with the elongating hypocotyls. After 10 days the roots of three uniform seedlings per genotype were imaged using a flatbed scanner (Epson® Expression 10000XL) and WinRhizo®. Images were analyzed using Automatic Root Image Analyzer (ARIA®) Version 2.0 (Pace et al. 2014) (Table 1). Shoot and root dry weights for each seedling were measured after a minimum dry down period of 72 hours at 75°C.

Data analysis: Mean, standard deviation (SD), variance, and range were calculated using JMP® Version 12 (SAS Institute Inc., Cary, NC, 1989-2017). ANOVA was used to investigate the effects of genotype and seed lot (i.e., year and replication within year) using PROC MIXED in SAS® Version 9.4. Pearson correlations were calculated using PROC CORR (SAS®). PCA analysis was conducted for seed and seedling root traits; z-scores (i.e., standardized scores) were calculated using Excel and analysis was performed in R 3.3.1 using prcomp( ). Genotypes were assigned to three populations (i.e., restorer lines (R), maintainer lines (B/HA), non-classified lines (N)) based on the results from the STRUCTURE and STRUCTURE HARVEST analyses described below. ANOVA was used to examine the significance of these
population assignments. Broad sense heritability ($H^2$) estimates were calculated using the method described in Pace et al. 2015.

Genotype data and genome wide association mapping: A total of 5,788 polymorphic (A/T) SNP markers were available for 271 genotypes (Bachlava et al. 2012). Population structure was estimated using STRUCTURE (Pritchard et al. 2000) according to Mandel et al. 2013. These results were then analyzed using STRUCTURE HARVESTER (Earl 2012), and the population number $K = 3$ was determined using the DeltaK method (Evanno, Regnaut and Goudet 2005). Best linear unbiased prediction (BLUP) was performed (R 3.3.1, LME4 package) (Supplemental Figure 4) prior to association mapping analyses. Population structure, kinship, and genome wide association mapping were conducted using GAPIT (Lipka et al. 2012). All 5,788 SNPs were used with a minor allele frequency set at $> 0.01$. Markers exceeding the Bonferroni correction threshold of $P > 8.63 \times 10^{-6}$ ($\alpha = 0.05/5,788$ markers) were declared significant. A false discovery rate (FDR) of $\leq 0.20$ was also used.

Results

Seed and seedling root phenotypes

Average germination time was 42.5 hours with most genotypes (i.e., $> 75\%$) germinating within 48.5 hours. Genotype was significant for germination time ($P < 0.0001$) as well as year ($P < 0.0001$) and replication within year ($P < 0.0009$). Average 100-seed weight was 5.85 g ($SD = 2.8$ g, range from 1.33 to 18.15 g). Average seed area was 45.5 mm$^2$ ($SD = 19.3$ mm$^2$, range from 16.13 to 118.96 mm$^2$). Average seed perimeter was 27.9 mm ($SD = 5$ mm); however, most genotypes (i.e., $> 75\%$) had perimeter values $< 31$ mm. Average seed length and width were 10.95 mm ($SD = 1.75$ mm) and 5.24 mm ($SD = 1.43$ mm), respectively. All
genotypes except for four genotypes with long narrow seeds had average circularity values within three standard deviations (i.e., SD = 0.06) of the population mean = 0.7.

Root length was highly variable; average total root length was 169 cm (SD = 81 cm) and average primary root length was 29 cm (SD = 7 cm) (Table 2). Total root length and primary root length were shorter ($P \leq 0.001$) for seed lots harvested in 2013 vs. 2014, but replications within each year were not statistically different. Seedling shoot and root weights were variable with shoot to root ratios ranging from 0.4 to 39.9 (i.e., 25% and 75% quantiles were 2.57 and 3.80, respectively).

Board-sense heritability ($H^2$) estimates (Table 2) were moderately high to high (0.41 - 0.88) for root traits, high (0.78 - 0.85) for seed size and shape traits, and moderate for seed weight and germination time (0.69 and 0.39, respectively).

Pearson correlations

Germination time was significantly correlated with seed traits (e.g., $r = 0.42$ and $0.55$ for seed length and width, respectively), but it was not correlated with primary root length ($r = 0.11$) (Table 3). Seed weight and seed size/shape traits were highly correlated ($r > 0.80$) except for seed circularity which was less correlated (i.e., $r = 0.63$). Total root length and secondary root length were moderately correlated ($r = 0.46 - 0.70$) with seed size/shape traits and only weakly correlated with germination time ($r = 0.18$ and 0.19, respectively). Primary root length was correlated ($r = 0.26 - 0.43$) with seed size/shape traits.

PCA analysis

PCA of seedling root traits explained 71.3% (PC1) and 18.9% (PC2) of the total variation (Figure 1). PCA of seed traits explained 78.1% (PC1) and 12.1% (PC2) of the total variation.
Restorer lines had shorter roots (P-value > 0.0001) and smaller seeds (P-value > 0.0001) relative to the maintainer lines and non-classified lines, which were not significantly different from each other (P-value = 0.11). The same results were found when HA vs. R-lines were compared within a usage group (i.e., just oil lines, just non-oil lines) (i.e., P-values < 0.001 - 0.011).

Marker and GWA

STRUCTURE and STRUCTURE HARVESTER both yielded K = 3, which was in agreement with the previous finding by Mandel et al. 2013. The 10-marker average of LD decay had several peaks between ~1800-4100 cM (Figure 2) indicating that the genome has several highly conserved regions with less recombination. The same was illustrated in Mandel et al. 2013, where chromosomes 5, 10, and 13 were found to contain large regions with high r² values.

Based on a FDR ≤ 0.20, 29 unique markers were associated with area, PL, W, L, PRL, RTW, STW, TTW and SDW (Table 4; 24 were identified using the 2013 data, 3 were identified using the 2014 data, and 2 were identified using the combined date set). Markers 3032 and 3552 were associated with primary root length and root dry weight, respectively, and were corroborated when seed weight or seed width were used as covariates in the BLUP model (Table 4 and Supplemental Figure 4). Several markers were associated with multiple traits: M1563 was associated with TTW, STW, and SDW and M3120 was associated with PL, W, and SDW. Four markers (i.e., M1563, M3032, M3036, and M3140) exceeded the more stringent Bonferroni correction threshold (P = 8.64 X 10⁻⁶). M1563 was located on chromosome 5 at ~43 cM and was significantly associated with STW (P-value = 1.91 X 10⁻⁶).
(Figure 3A). M3032 and M3036 were located on chromosome 9 at ~79 cM and ~81 cM and were both associated with PRL (Figure 3B). The QQ plot for PRL (Figure 3C) contained one small peak and one large peak, which suggested that there are two QTL in this narrow region. M3140 was located on chromosome 10 at ~28 cM and was associated with seed width (Figure 3C). There were several markers in proximity to M3120 and M3032.

Discussion

Based on our knowledge, this experiment was the first to characterize seedling root traits in a diverse panel of sunflower genotypes. The variation in both seed and seedling root traits was quite high and was suitably captured using 2D digital images acquired with a flatbed scanner. Seedling root growth for seeds produced in different years (i.e., 2013 vs. 2014) was different, thus indicating that early seedling root growth was likely affected by the environment in which the seed was produced. This observation was not unexpected, because seed quality and seed origin have been shown in other studies to influence multiple phenotypes including seed germination rate (Ahmad 2001). Therefore, utilizing seed from multiple seed lots is recommended when studying seedling root traits to fully understand the genetic variation that exists.

In this study seed weight was positively correlated with total root length (r = 0.70) and total seedling dry weight and root dry weight (r = 0.90 and 0.74, respectively). A similar correlation between seed weight and seedling root length was observed in wheat, lovegrass (e.g., Eragrostis barrelieri, Eragrostis dielsii, etc.) and soybean (Marco and De Marco 1990, Westoby, Jurado and Leishman 1992, Lynch and van Beem 1993). This suggests that seed weight (i.e., the size of the cotyledons in a sunflower seed) may affect seedling root growth
such that seeds containing more stored energy (i.e., heavier seeds) produce bigger seedlings with longer roots. In comparison, total root length was less correlated with seed size ($r = 0.56$ vs $0.70$) despite the high correlation between seed weight and seed area ($r = 0.92$) indicating that larger sunflower seeds may not always contain larger cotyledons to support increased root growth, but may instead have larger, thicker hulls (Prasifka, Hulke and Seiler 2014).

Variation in the seedling shoot to root ratios observed in this study suggests that the energy stored in sunflower seeds (i.e., cotyledons) may be partitioned differently between shoots and roots during the first ten days of growth resulting in some genotypes having lower shoot to root weight ratios (i.e., $< 2.8$) as compared to the average of 3.3. If in fact there is genetic variation in the rate of energy partitioning in sunflower seedlings in roots vs. shoots, then this could be selected for to improve emergence and seedling establishment in hybrids (Somers, Ullrich and Ramsay 1983, Hussain et al. 2006, Ahmad et al. 2009).

Broad sense heritability ($H^2$) estimates from this study for root traits were between $0.77 - 0.88$. These estimates were higher than or comparable to those observed in other experiments [i.e., $0.43$ in maize (Pace et al. 2015), $0.63$ in Arabidopsis (Loudet et al. 2005), $0.70$ in wheat (Monyo and Whittington 1970), and $0.69 - 0.90$ in pea (O'Toole and Bland 1987)]. The high values observed in this study were likely attributable to the reduction (or elimination) of genotype by environment interactions since the conditions in the growth chambers were close to optimal and highly consistent. The heritability ($H^2$) estimates from
this study for seed size (e.g., $H^2 = 0.85$ for area, width, and length) were very similar to the narrow sense heritability ($h^2$) estimates ($> 0.90$) observed in Tang et al. 2006.

Principle component analysis and ANOVA indicated that seed and seedling root traits were distinct in the two heterotic groups in sunflower with restorer lines (R-lines) having smaller seeds and shorter roots relative to maintainer lines (B/HA-lines) (Figure 1). The same results were observed within usage classes (i.e., oil R-lines vs. oil B/HA lines and non-oil R-lines vs. non-oil B/HA lines). The difference between heterotic groups for seedling root traits could be indirectly attributed to selection by breeders for apical branching in R-lines vs. no branching (monocephaly) in B/HA-lines if selection for branching in shoots somehow influences genes that contribute to branching in roots. However, because the impact of above ground selection (i.e., domestication/breeding) on root architecture is not well understood (Waines and Ehdaie 2007, Burton, Brown and Lynch 2013) this hypothesis will need further examination.

Chromosome 5 in sunflower harbors QTL for flowering time, seed weight, and plant height (León, Lee and Andrade 2001, Bert et al. 2003, Tang et al. 2006, Nambeesan et al. 2015). In this study, we identified a QTL on Chromosome 5 for seedling shoot weight located near a QTL for seed length identified in Tang et al. 2006. Because seed length was highly correlated ($r = 0.80$) with shoot weight in this study, it is possible that a single QTL is either directly or indirectly responsible for both traits (e.g., Song et al. 2007) though additional experimentation is needed to determine the validity of this hypothesis.

Chromosome 10 in sunflower contains several highly conserved regions, which contain QTL for traits (i.e., plant height, flowering time, photosynthesis, 100-seed weight, seed oil
content, etc.) (Tang et al. 2006, Kiani et al. 2007, Nambeesan et al. 2015) that are under intense selection by breeders because they directly affect overall fitness and yield of a plant. Marker 3120 on chromosome 10 was associated with seed weight, width and perimeter and was located near a previously identified QTL for oil content (Ebrahimi et al. 2008). Assuming that the negative correlation between oil content and seed size identified in soybean and other sunflower populations (Fick and Miller 1997, Fick, Zimmer and Zimmerman 1974, Panthee et al. 2005) exists in this panel it is likely that oil content data from this population would lead to the identification of a QTL for oil content in this same region.

Conclusion

This study demonstrated that seedling root architecture in sunflower is highly variable, and that 2D digital images acquired with a flatbed scanner and analyzed with ARIA® were suitable for capturing this variation. Markers associated with seedling root and seed traits identified by this study could be used to identify candidate genes that could be used by breeders for MAS after additional testing and verification is completed.
### Table 1. Seed and root trait designations and descriptions

<table>
<thead>
<tr>
<th>Trait name</th>
<th>Abbreviation</th>
<th>Trait description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Germination time</td>
<td>GT</td>
<td>Time required to germinate 10 seeds (hours)</td>
</tr>
<tr>
<td>100-seed weight</td>
<td>SDW</td>
<td>100-seed weight (g)</td>
</tr>
<tr>
<td>Seed area*</td>
<td>Area</td>
<td>Seed area (mm²)</td>
</tr>
<tr>
<td>Seed perimeter*</td>
<td>PL</td>
<td>Seed perimeter (mm)</td>
</tr>
<tr>
<td>Seed length*</td>
<td>L</td>
<td>Seed length (mm)</td>
</tr>
<tr>
<td>Seed width*</td>
<td>W</td>
<td>Seed width (mm)</td>
</tr>
<tr>
<td>Seed circularity*</td>
<td>CIR</td>
<td>Seed circularity</td>
</tr>
<tr>
<td>Total root length**</td>
<td>TRL</td>
<td>Total length of all roots (cm)</td>
</tr>
<tr>
<td>Total secondary root length**</td>
<td>SRL</td>
<td>Total length of all secondary roots (cm)</td>
</tr>
<tr>
<td>Primary root length**</td>
<td>PRL</td>
<td>Length of the primary root (cm)</td>
</tr>
<tr>
<td>Shoot weight</td>
<td>STW</td>
<td>Shoot dry weight (g)</td>
</tr>
<tr>
<td>Root weight</td>
<td>RTW</td>
<td>Root dry weight (g)</td>
</tr>
<tr>
<td>Total seedling weight</td>
<td>TTW</td>
<td>The sum of shoot dry weight + root dry weight (g)</td>
</tr>
<tr>
<td>STW/RTW ratio</td>
<td>SRR</td>
<td>The ratio of shoot dry weight to root dry weight</td>
</tr>
</tbody>
</table>

*traits captured by Smartgrain® Version 1.1, **traits captured using ARIA® Version 2.0.
Table 2. Seed and seedling root traits from seed produced in 2013 and 2014

<table>
<thead>
<tr>
<th>Traits</th>
<th>2013</th>
<th></th>
<th>2014</th>
<th></th>
<th>H²</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Maximum</td>
<td>Minimum</td>
<td>Mean</td>
</tr>
<tr>
<td>GT (h)</td>
<td>44.44</td>
<td>9.31</td>
<td>98.59</td>
<td>26.40</td>
<td>40.66</td>
</tr>
<tr>
<td>SDW (g)</td>
<td>6.38</td>
<td>3.10</td>
<td>18.15</td>
<td>1.73</td>
<td>5.33</td>
</tr>
<tr>
<td>Area (mm²)</td>
<td>46.10</td>
<td>19.90</td>
<td>118.96</td>
<td>16.18</td>
<td>44.89</td>
</tr>
<tr>
<td>PL (mm)</td>
<td>28.18</td>
<td>5.19</td>
<td>43.82</td>
<td>17.18</td>
<td>27.55</td>
</tr>
<tr>
<td>L (mm)</td>
<td>11.10</td>
<td>1.80</td>
<td>16.51</td>
<td>6.95</td>
<td>10.80</td>
</tr>
<tr>
<td>W (mm)</td>
<td>5.22</td>
<td>1.46</td>
<td>10.16</td>
<td>2.60</td>
<td>5.25</td>
</tr>
<tr>
<td>CIR</td>
<td>0.69</td>
<td>0.07</td>
<td>0.84</td>
<td>0.45</td>
<td>0.71</td>
</tr>
<tr>
<td>TRL (cm)</td>
<td>161.64</td>
<td>78.36</td>
<td>536.72</td>
<td>12.43</td>
<td>178.00</td>
</tr>
<tr>
<td>SRL (cm)</td>
<td>132.45</td>
<td>74.40</td>
<td>485.12</td>
<td>0.00</td>
<td>148.25</td>
</tr>
<tr>
<td>PRL (cm)</td>
<td>29.20</td>
<td>7.13</td>
<td>54.10</td>
<td>5.04</td>
<td>29.76</td>
</tr>
<tr>
<td>STW (g)</td>
<td>0.0321</td>
<td>0.0138</td>
<td>0.0848</td>
<td>0.0068</td>
<td>0.0301</td>
</tr>
<tr>
<td>RTW (g)</td>
<td>0.0104</td>
<td>0.0048</td>
<td>0.0371</td>
<td>0.0003</td>
<td>0.0098</td>
</tr>
<tr>
<td>TTW (g)</td>
<td>0.0425</td>
<td>0.0177</td>
<td>0.1153</td>
<td>0.0071</td>
<td>0.0399</td>
</tr>
<tr>
<td>SSR</td>
<td>3.31</td>
<td>1.42</td>
<td>35.50</td>
<td>0.58</td>
<td>3.35</td>
</tr>
</tbody>
</table>

Pl = seed perimeter, L = length, W = width, CIR = circularity, STW = shoot weight, RTW = root weight, TTW = STW + RTW, SSR = shoot to root ratio, SDW = seed weight, GT = germination time, TRL = total root length, PRL = primary root length, SRL = secondary root length
### Table 3. Pearson (r) correlations

<table>
<thead>
<tr>
<th></th>
<th>SDW</th>
<th>GT</th>
<th>Area</th>
<th>PL</th>
<th>L</th>
<th>W</th>
<th>CIR</th>
<th>RTW</th>
<th>STW</th>
<th>TTW</th>
<th>SSR</th>
<th>TRL</th>
<th>SRL</th>
<th>PRL</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDW</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GT</td>
<td>0.49</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area</td>
<td>0.93</td>
<td>0.51</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PL</td>
<td>0.92</td>
<td>0.48</td>
<td>0.98</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>0.87</td>
<td>0.42</td>
<td>0.92</td>
<td>0.97</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W</td>
<td>0.92</td>
<td>0.55</td>
<td>0.96</td>
<td>0.92</td>
<td>0.82</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CIR</td>
<td>0.63</td>
<td>0.50</td>
<td>0.61</td>
<td>0.51</td>
<td>0.35</td>
<td>0.78</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RTW</td>
<td>0.74</td>
<td>0.18</td>
<td>0.57</td>
<td>0.61</td>
<td>0.60</td>
<td>0.59</td>
<td>0.41</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STW</td>
<td>0.92</td>
<td>0.39</td>
<td>0.82</td>
<td>0.83</td>
<td>0.80</td>
<td>0.83</td>
<td>0.58</td>
<td>0.85</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TTW</td>
<td>0.90</td>
<td>0.35</td>
<td>0.78</td>
<td>0.80</td>
<td>0.77</td>
<td>0.79</td>
<td>0.55</td>
<td>0.91</td>
<td>0.99</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSR</td>
<td>0.32</td>
<td>0.35</td>
<td>0.40</td>
<td>0.37</td>
<td>0.33</td>
<td>0.39</td>
<td>0.28</td>
<td>-0.20</td>
<td>0.27</td>
<td>0.16</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRL</td>
<td>0.70</td>
<td>0.18</td>
<td>0.56</td>
<td>0.58</td>
<td>0.54</td>
<td>0.60</td>
<td>0.46</td>
<td>0.90</td>
<td>0.79</td>
<td>0.84</td>
<td>-0.16</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRL</td>
<td>0.71</td>
<td>0.19</td>
<td>0.57</td>
<td>0.58</td>
<td>0.55</td>
<td>0.60</td>
<td>0.46</td>
<td>0.90</td>
<td>0.79</td>
<td>0.84</td>
<td>-0.16</td>
<td>1.00</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>PRL</td>
<td>0.43</td>
<td>0.11</td>
<td>0.29</td>
<td>0.31</td>
<td>0.29</td>
<td>0.33</td>
<td>0.26</td>
<td>0.59</td>
<td>0.51</td>
<td>0.55</td>
<td>-0.17</td>
<td>0.70</td>
<td>0.65</td>
<td>1</td>
</tr>
</tbody>
</table>

PL = seed perimeter, L = length, W = width, CIR = circularity, STW = shoot weight, RTW = root weight, TTW = STW + RTW, SSR = shoot to root ratio, SDW = seed weight, GT = germination time, TRL = total root length, PRL = primary root length, SRL = secondary root length
<table>
<thead>
<tr>
<th>Marker</th>
<th>Chromosome</th>
<th>Position</th>
<th>Trait</th>
<th>Data set</th>
<th>FDR</th>
<th>Bonferroni</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1563</td>
<td>5</td>
<td>43423</td>
<td>SDW, STW, TTW</td>
<td>all, 2014</td>
<td>0.0109</td>
<td>Yes</td>
</tr>
<tr>
<td>M1760</td>
<td>5</td>
<td>73171</td>
<td>Area</td>
<td>2014</td>
<td>0.1597</td>
<td></td>
</tr>
<tr>
<td>M2520</td>
<td>8</td>
<td>45752</td>
<td>W</td>
<td>2013</td>
<td>0.0640</td>
<td></td>
</tr>
<tr>
<td>M2529</td>
<td>8</td>
<td>45931</td>
<td>SDW</td>
<td>all</td>
<td>0.1663</td>
<td></td>
</tr>
<tr>
<td>M2542</td>
<td>8</td>
<td>46656</td>
<td>L</td>
<td>2013</td>
<td>0.1526</td>
<td></td>
</tr>
<tr>
<td>M2544</td>
<td>8</td>
<td>46658</td>
<td>Area</td>
<td>2014</td>
<td>0.1597</td>
<td></td>
</tr>
<tr>
<td>M3029</td>
<td>9</td>
<td>79101</td>
<td>PRL</td>
<td>2013, 2013s*</td>
<td>0.0948</td>
<td></td>
</tr>
<tr>
<td>M3030</td>
<td>9</td>
<td>79810</td>
<td>PRL</td>
<td>2013</td>
<td>0.0327</td>
<td></td>
</tr>
<tr>
<td>M3031</td>
<td>9</td>
<td>79900</td>
<td>PRL</td>
<td>2013</td>
<td>0.0312</td>
<td></td>
</tr>
<tr>
<td>M3032</td>
<td>9</td>
<td>79901</td>
<td>PRL</td>
<td>2013, 2013s, 2013w</td>
<td>0.0204</td>
<td>Yes</td>
</tr>
<tr>
<td>M3033</td>
<td>9</td>
<td>79902</td>
<td>PRL</td>
<td>2013</td>
<td>0.0919</td>
<td></td>
</tr>
<tr>
<td>M3036</td>
<td>9</td>
<td>81470</td>
<td>PRL</td>
<td>2013, 2013s, 2013w</td>
<td>0.0204</td>
<td>Yes</td>
</tr>
<tr>
<td>M3037</td>
<td>9</td>
<td>81500</td>
<td>PRL</td>
<td>2013, 2013s</td>
<td>0.0312</td>
<td></td>
</tr>
<tr>
<td>M3108</td>
<td>10</td>
<td>20130</td>
<td>W</td>
<td>2013</td>
<td>0.1146</td>
<td></td>
</tr>
<tr>
<td>M3118</td>
<td>10</td>
<td>24990</td>
<td>W</td>
<td>all, 2013</td>
<td>0.0539</td>
<td></td>
</tr>
<tr>
<td>M3120</td>
<td>10</td>
<td>25610</td>
<td>SDW, W, PL</td>
<td>all, 2013</td>
<td>0.0142</td>
<td>Yes</td>
</tr>
<tr>
<td>M3123</td>
<td>10</td>
<td>26022</td>
<td>W</td>
<td>2013</td>
<td>0.1016</td>
<td></td>
</tr>
<tr>
<td>M3126</td>
<td>10</td>
<td>26421</td>
<td>W</td>
<td>all, 2013</td>
<td>0.0539</td>
<td></td>
</tr>
<tr>
<td>M3131</td>
<td>10</td>
<td>27832</td>
<td>SDW, W, PL</td>
<td>all, 2013</td>
<td>0.0539</td>
<td></td>
</tr>
<tr>
<td>M3132</td>
<td>10</td>
<td>27890</td>
<td>W, PL</td>
<td>all, 2013</td>
<td>0.0539</td>
<td></td>
</tr>
<tr>
<td>M3140</td>
<td>10</td>
<td>28910</td>
<td>L, W, PL</td>
<td>2013</td>
<td>0.0430</td>
<td></td>
</tr>
<tr>
<td>M3143</td>
<td>10</td>
<td>28913</td>
<td>L, W, PL</td>
<td>2013</td>
<td>0.0430</td>
<td></td>
</tr>
<tr>
<td>M3144</td>
<td>10</td>
<td>29000</td>
<td>L, W, PL</td>
<td>2013</td>
<td>0.0430</td>
<td></td>
</tr>
<tr>
<td>M3145</td>
<td>10</td>
<td>29230</td>
<td>L, W, PL</td>
<td>2013</td>
<td>0.0430</td>
<td></td>
</tr>
<tr>
<td>M3146</td>
<td>10</td>
<td>29440</td>
<td>L, W, PL</td>
<td>2013</td>
<td>0.0430</td>
<td></td>
</tr>
<tr>
<td>M338</td>
<td>1</td>
<td>48102</td>
<td>W</td>
<td>2013</td>
<td>0.0640</td>
<td></td>
</tr>
<tr>
<td>M3552</td>
<td>10</td>
<td>61321</td>
<td>RTW</td>
<td>2014s, 2014w</td>
<td>0.0966</td>
<td></td>
</tr>
<tr>
<td>M5199</td>
<td>16</td>
<td>44722</td>
<td>PL</td>
<td>2013</td>
<td>0.2000</td>
<td></td>
</tr>
<tr>
<td>M5599</td>
<td>17</td>
<td>44117</td>
<td>W</td>
<td>2013</td>
<td>0.1694</td>
<td></td>
</tr>
</tbody>
</table>

*2013s and 2014s refers to the BLUP datasets using seed weight as covariate; 2013w and 2014w refers to the BLUP datasets using seed weight as covariate.

SDW = seed weight, STW = shoot weight, TTW = STW + RTW, W = width, L = length, PRL = primary root length, PL = seed perimeter, RTW = root length, FDR = false discovery rate.
Figure 1. PCA plot of root traits.
Figure 2. LD decay between two markers (open circles), average LD decay of 10 markers (red circles) (A). Marker density, the curve represents the cumulative frequency (B).
Figure 3. Manhattan plots and their corresponding QQ-plots for shoot weight (A), primary root length (B), and seed width (C). Green lines indicate the FDR thresholds, red lines represent the Bonferroni correction thresholds.
Supplemental Figure 1. Average annual precipitation in the US.
Supplemental Figure 2. PCA plot of selected seed and seedling root traits.
Supplemental Figure 3. PCA plot of seed traits.
Equation 1. $Y = \mu + Exp + Box(Exp) + Year + Rep(Year) + Geno + \epsilon$

Equation 2. $Y = \mu + Exp + Box(Exp) + Year + Rep(Year) + Geno + Trait + \epsilon$

<table>
<thead>
<tr>
<th>$Y$</th>
<th>The root trait being analyzed</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>The mean</td>
</tr>
<tr>
<td>$Exp$</td>
<td>The experiments (total $4 \times 2 = 8$)</td>
</tr>
<tr>
<td>$Box(Exp)$</td>
<td>The blocks within each experiment (total 6)</td>
</tr>
<tr>
<td>$Year$</td>
<td>The seed source 2013 and 2014 (total 2)</td>
</tr>
<tr>
<td>$Year(Rep)$</td>
<td>The seed source replication within year, 1 and 2 (total 2)</td>
</tr>
<tr>
<td>$Geno$</td>
<td>The genotype (288)</td>
</tr>
<tr>
<td>$E$</td>
<td>And the residual</td>
</tr>
</tbody>
</table>

Equation 3. $Y = \mu + Year + Rep(Year) + Geno + Year \times Geno + \epsilon$

<table>
<thead>
<tr>
<th>$Y$</th>
<th>The seed trait being analyzed</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>The mean</td>
</tr>
<tr>
<td>$Year$</td>
<td>The seed source 2013 and 2014 (total 2)</td>
</tr>
<tr>
<td>$Year(Rep)$</td>
<td>The seed source replication within year, 1 and 2 (total 2)</td>
</tr>
<tr>
<td>$Geno$</td>
<td>The genotype (288)</td>
</tr>
<tr>
<td>$Year*Geno$</td>
<td>The interaction between year and genotype</td>
</tr>
<tr>
<td>$E$</td>
<td>And the residual</td>
</tr>
</tbody>
</table>

Supplemental Figure 4. Equations 1 and 2 were used to analyze root traits, Equation 2 included seed traits (i.e., seed width and seed weight) as covariates, Equation 3 was used to analyze seed traits.


Weaver, J. E. & W. E. Bruner (1926) Root development of field crops.


