BMC
Genomics

Genome-wide association analysis of seedling root development in maize (*Zea mays L.*)

Jordon Pace[1*], Candice Gardner[1], Cinta Romay[2], Baskar Ganapathysubramanian[3] and Thomas Lübberstedt[1]

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 ($p \leq 5.3 \times 10^{-7}$). Analysis of significant SNP markers for multiple traits showed that several were located within gene models with some SNP markers localized within regions of 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.

Keywords: Genome wide association study, Maize, Roots

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 transmembrane 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

* Correspondence: jmpace1@iastate.edu
[1]Department of Agronomy, Iowa State University, Ames, Iowa 50013, USA
Full list of author information is available at the end of the article

Pace *et al. BMC Genomics* (2015) 16:47

Page 2 of 12

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 (H^2) 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

Pace *et al. BMC Genomics* (2015) 16:47

Page 3 of 12

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

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

Pace *et al. BMC Genomics* (2015) 16:47

Page 4 of 12

Table 2 Trait statistics collected for all 22 traits

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	0.049 g	0.019	.005 g	0.124 g	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.

at P = 5.36 × 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.
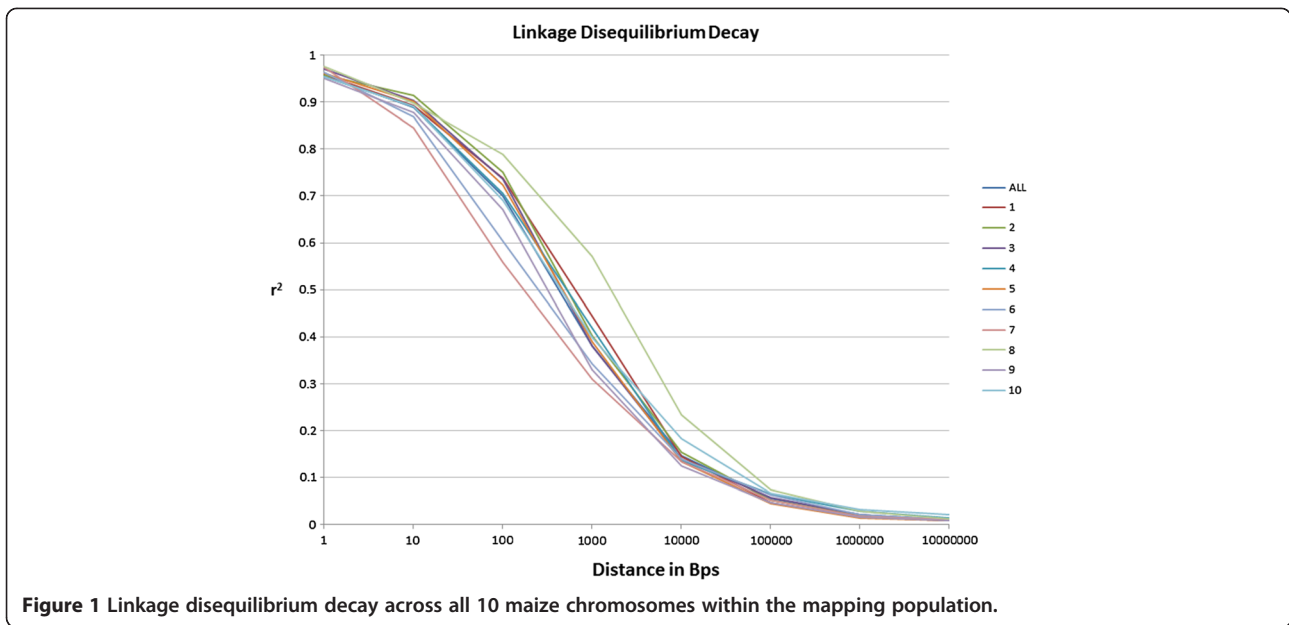
Using the GLM model, an additional 263 significant markers were found using the same threshold of P = 5.36 × 10-7 for root traits above the heritability threshold of H^2 ≥ 0.30. Clustering of significant SNPs using GLM was analyzed. SNPs associated with root traits clustered on chromosomes 2, 3, 4, and 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

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.

Pace *et al. BMC Genomics* (2015) 16:47

Page 6 of 12

Figure 1 Linkage disequilibrium decay across all 10 maize chromosomes within the mapping population.

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.
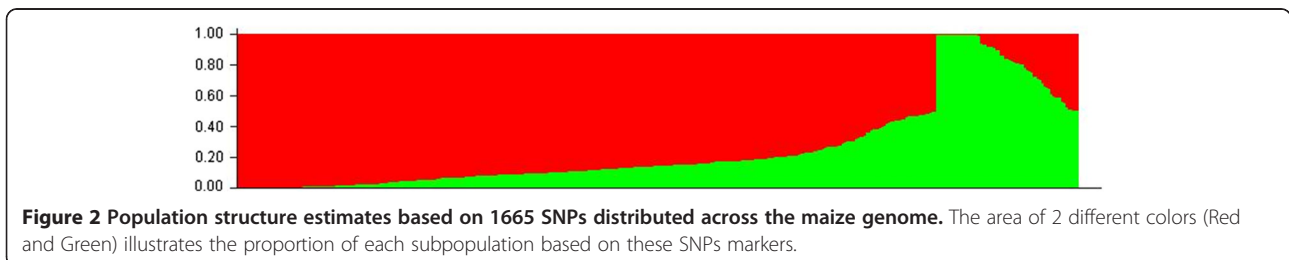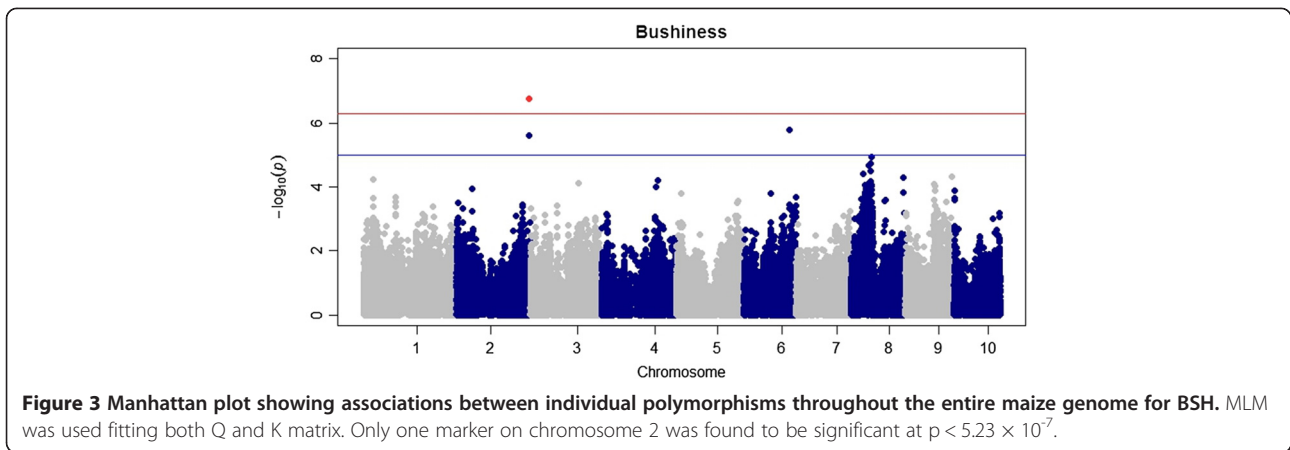
Discussion

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 heritabilities 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))

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 throughout 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 $p < 5.23 \times 10^{-7}$.

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

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 $p < 5.23 \times 10^{-7}$.

Pace *et al. BMC Genomics* (2015) 16:47

Page 8 of 12
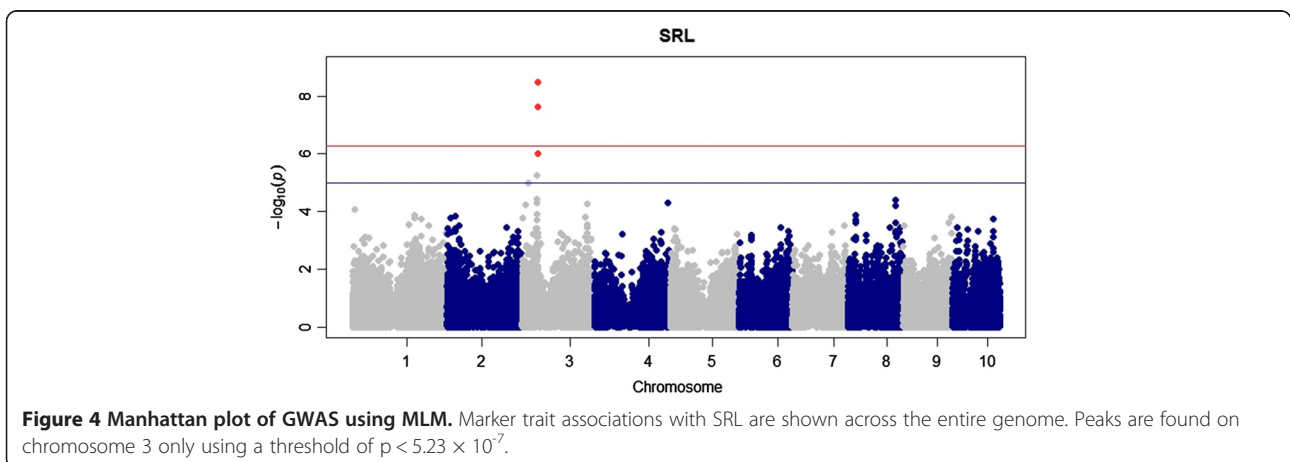
Table 4 Gene model absolute expression values found in B73 genome

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

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, 263 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. This gene model is located on chromosome 4 and contained 9 of 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

Pace *et al. BMC Genomics* (2015) 16:47

Page 9 of 12

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 (Additional file 4: Table S2). 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-moistened with fungicide solution Captan® (2.5 g/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°C. Light intensity was 200 µmol photons m-2 s-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

Pace *et al. BMC Genomics* (2015) 16:47

Page 10 of 12

conducted, roots and shoots were collected separately and dried for 48 hrs at 55°C, to determine RDW, SDW, and TPB. In addition, 22 traits (Table 1 and Additional file 1: Figure S1) 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 $\left(\sigma_g^2\right)$, and phenotypic $\left(\sigma_p^2\right)$ 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:

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

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

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

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).

Pace *et al. BMC Genomics* (2015) 16:47

Page 11 of 12

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 1 s and 0 s, 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 $G = 2K\sigma^2 G$. 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 5.3×10^{-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 ($\alpha/(Meff\_G)$). Markers above the suggested threshold for MLM were considered as significantly trait-associated SNP markers and candidates for causative polymorphisms.

Availability of supporting data

Phenotypic data will be available upon request from the reader. Genotypic data can be found freely available at http:\\www.panzea.org/lit/data_sets.html. As the GBS data used for this study is a subset of the entire GBS Ames US Inbreds data set.

Additional files

Additional file 1: Figure S1. Illustrations of the parameters measured by *ARIA* for seedling root traits extracted for GWAS.

Additional file 2: Table S1. List of all significant marker trait associations determined by GWAS using MLM and GLM.

Additional file 3: Figure S2. Manhattan plots showing associations for traits using GLM association analysis. Dots above the red line indicate putatively associated SNPs with each trait.

Additional file 4: Table S2. All lines included in 384 Ames Panel Association mapping population.

Abbreviations

GWAS: Genome wide association study; LD: Linkage disequilibrium; GLM: General Linear Model; MLM: Mixed linear model; GBS: Genotype by sequencing; NUE: Nitrogen use efficiency, Table 2 gives all trait abbreviations.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

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. All authors read and approved the final manuscript.

Author details

[1]Department of Agronomy, Iowa State University, Ames, Iowa 50013, USA. [2]Institute for Genomic Diversity, Cornell University, Ithaca, NY 14853, USA. [3]Department of Mechanical Engineering, Iowa State University, Ames, Iowa 50013, USA.

References

1. FAO. Current world fertilizer trends and outlook to 2014. Rome: Food and Agriculture Organization of the United Nations; 2010.
2. Tilman D, Cassman KG, Matson P, Naylor R, Polasky S. Agriculture sustainability and intensive production practices. Nature. 2002;418:671–7.
3. Raun WR, Johnson GV. Improving nitrogen use efficiency for cereal production. Agron J. 1999;91(3):357–63.
4. Quemada LM. Strategies to improve nitrogen use efficiency in winter cereal crops under rainfed conditions. Agron J. 2008;100(2):277–84.
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. J Plant Nutr. 2006;29(2):317–30.
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):2369–87.
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. Mol Breed. 2007;20(1):15–29.
8. Hodge A, Berta G, Doussan C, Merchan F, Crespi M. Plant root growth, architecture and function. Plant Soil. 2009;321(1–2):153–87.
9. Hochholdinger F. The maize root system: morphology, anatomy, and genetics. In: Bennetzen JL, Hake SC, editors. Handbook of Maize: Its Biology. 2009.
10. Hoppe DC, McCully ME, Wenzel CL. The nodal roots of Zea: their development in relation to structural features of the stem. Can J Bot. 1986;64(11):2524–37.
11. Lynch J. Root Architecture and Plant Productivity. Plant Physiol. 1995;109:7–13.
12. Mackay S. Root Growth and phosphorus and potassium uptake by two corn genotypes in the field. Fertilizer Res. 1986;10:217–30.
13. Marschener H. Role of root growth, arbuscular mycorrhiza, and root exudates for the efficiency in nutrient acquisition. Field Crops Res. 1998;56(1–2):203–7.

Pace *et al. BMC Genomics* (2015) 16:47

Page 12 of 12

14. Horst WJ, Kamh M, Jibrin JM, Chude VO. Agronomic measures for increasing P availability to crops. Plant Soil. 2001;237:211–23.
15. Taramino G, Sauer M, Stauffer JL, Multani D, Niu X, Sakai H, et al. 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):649–59.
16. Hochholdinger F, Wen TJ, Zimmermann R, Chimot-Marolle P, da Costa e Silva O, Bruce W, et al. 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):888–98.
17. Yan J, Warburton M, Crouch J. Association Mapping for Enhancing Maize (L.) Genetic Improvement. Crop Sci. 2011;51(2):433.
18. Buckler ES, Thornsberry JM. Plant molecular diversity and applications to genomics. Curr Opin Plant Biol. 2002;5:107–11.
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):155–60.
21. Risch N, Merkangas K. The future of genetic studies of complex human diseases. Science. 1996;273(5281):1516–7.
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):e56495.
24. Romay MC, Millard MJ, Glaubitz JC, Peiffer JA, Swarts KL, Casstevens TM, 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, et al. 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):123–33.
26. Sekhon RS, Lin H, Childs KL, Hansey CN, Buell CR, de Leon N, et al. Genome-wide atlas of transcription during maize development. Plant J. 2011;66(4):553–63.
27. Peiffer JA, Romay MC, Gore MA, Flint-Garcia SA, Zhang Z, Millard MJ, et al. The Genetic Architecture of Maize Height. Genetics. 2014;196(4):1337–56.
28. Cai H, Chen F, Mi G, Zhang F, Maurer HP, Liu W, et al. Mapping QTLs for root system architecture of maize (Zea mays L.) in the field at different developmental stages. Theor Appl Genet. 2012;125(6):1313–24.
29. Pritchard JK, Stephens M, Donnelly P. Inference of Population Structure Using Multilocus Genotype Data. Genetics. 2000;155:945–59.
30. Wu X, Li Y, Shi Y, Song Y, Wang T, Huang Y, et al. Fine genetic characterization of elite maize germplasm using high-throughput SNP genotyping. Theor Appl Genet. 2014;127(3):621–31.
31. Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, et al. The genetic architecture of maize flowering time. Science. 2009;325(5941):714–8.
32. Brown PJ, Upadyayula N, Mahone GS, Tian F, Bradbury PJ, Myles S, 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, et al. 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 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. Theor Appl Genet. 2005;111(4):688–95.
36. Xue Y, Warburton ML, Sawkins M, Zhang X, Setter T, Xu Y, et al. Genome-wide association analysis for nine agronomic traits in maize under well-watered and water-stressed conditions. Theor Appl Genet. 2013;126(10):2587–96.
37. Hansey CN, Vaillancourt B, Sekhon RS, De Leon N, Kaeppler SM, Buell CR. Maize (Zea mays L.) Genome Diversity as Revealed by RNA-Sequencing. PloS One. Genetics. 2012;7(3):e33071.
38. Kusano H, Testerink C, Vermeer JEM, Tsuge T, Shimada H, Oka A, et al. The Arabidopsis Phosphatidylinositol Phosphate 5-Kinase PIP5K3 Is a Key Regulator of Root Hair Tip Growth. Plant Cell. 2008;20:367–80.
39. Zhang Z, Ersoz E, Lai CQ, Todhunter RJ, Tiwari HK, Gore MA, et al. Mixed linear model approach adapted for genome-wide association studies. Nat Genet. 2010;42(4):355–60.
40. Lynch JP. Steep, cheap and deep: an ideotype to optimize water and N acquisition by maize root systems. Ann Bot. 2013;112(2):347–57.
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):1369–73.
43. Postma JA, Dathe A, Lynch J. The optimal lateral root branching density for maize depends on nitrogen and phosphorus availability. Plant Physiol. 2014;166(2):590–602.
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):e108255.
45. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. 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):2611–20.
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):2633–5.
48. Hardy OJ, Vekemans X. SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Mol Ecol Notes. 2002;2:618–20.
49. Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, et al. GAPIT: genome association and prediction integrated tool. Bioinformatics. 2012;28(18):2397–9.
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):361–9.
51. Johnson RC, Nelson GW, Troyer JL, Lautenberger JA, Kessing BD, Winkler CA, et al. Accounting for multiple comparisons in a genome-wide association study (GWAS). BMC Genomics. 2010;11:724.