Iowa State University
Digital Repository

Agronomy Publications | Agronomy

7-2014

Association analysis of single nucleotide polymorphisms in candidate genes with root traits in maize (Zea mays L.) seedlings

Bharath Kumar
Iowa State University

Adel H. Abdel-Ghani
Mu' tah University

Jordan Pace
Iowa State University

Jenaro Reyes-Matamors
Autonomous University of Puebla

Frank Hochholdinger
University of Bonn

See next page for additional authors

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

Part of the Agricultural Science Commons, Agronomy and Crop Sciences Commons, Genetics and Genomics Commons, and the Plant Breeding and Genetics Commons

The complete bibliographic information for this item can be found at https://lib.dr.iastate.edu/agron_pubs/455. For information on how to cite this item, please visit http://lib.dr.iastate.edu/howtocite.html.

This Article is brought to you for free and open access by the Agronomy at Iowa State University Digital Repository. It has been accepted for inclusion in Agronomy Publications by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

Association analysis of single nucleotide polymorphisms in candidate genes with root traits in maize (Zea mays L.) seedlings

Abstract

Background: Root growth and development is not only critical for nitrogen acquisition in plants, but also to anchor the plant in the soil. Several genes involved in maize root development have been isolated. Identification of SNPs associated with root traits would enable the selection of maize lines with better root architecture that might help to improve N uptake, and consequently plant growth particularly under N deficient conditions.

Results: In the present study, an association study (AS) panel consisting of 74 maize inbred lines was screened for seedling root traits in 6, 10, and 14-day-old seedlings. Allele resequencing of candidate root genes Rtcl, Rth3, Rum1, and Rul1 was also carried out in the same AS panel lines. All four candidate genes displayed different levels of nucleotide diversity, haplotype diversity and linkage disequilibrium. Nucleotide diversity was highest in the Rtcl gene (π=0.021), intermediate in Rum1 (π=0.011), lowest in Rth3 (π=0.007) and Rul1 (π=0.005) gene. When coding and non-coding regions within the genes were compared, nucleotide diversity varied across the genes. Gene based association analyses were carried out between individual polymorphisms in candidate genes, and root traits measured in 6, 10, and 14-day-old maize seedlings. Association analyses revealed several polymorphisms within the Rtcl, Rth3, Rum1, and Rul1 genes associated with seedling root traits. These significantly associated SNPs also affected putative functional sequence motifs, mostly transcription factor binding sites, and major domains in the genes.

Conclusion: Several nucleotide polymorphisms in Rtcl, Rth3, Rum1, and Rul1 were significantly ($P<0.05$) associated with seedling root traits in maize suggesting that all four tested genes are involved in the maize root development. Thus considerable allelic variation present in these root genes can be exploited for improving maize root characteristics. Target nucleotide polymorphisms for functional marker development were identified which might find application in marker-based selection strategies in breeding programs.

Keywords

Maize, Root traits, Single nucleotide polymorphism, Nucleotide diversity, Linkage disequilibrium, Gene based association mapping

Disciplines

Agricultural Science | Agronomy and Crop Sciences | Genetics and Genomics | Plant Breeding and Genetics

Comments

This is a manuscript of an article published as Kumar, Bharath, Adel H. Abdel-Ghani, Jordon Pace, Jenaro Reyes-Matamoros, Frank Hochholdinger, and Thomas Lübberstedt. "Association analysis of single nucleotide polymorphisms in candidate genes with root traits in maize (Zea mays L.) seedlings." *Plant science* 224 (2014): 9-19. doi: 10.1016/j.plantsci.2014.03.019. Posted with permission.

This article is available at Iowa State University Digital Repository: https://lib.dr.iastate.edu/agron_pubs/455

Creative Commons License

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License.

Authors
Bharath Kumar, Adel H. Abdel-Ghani, Jordan Pace, Jenaro Reyes-Matamors, Frank Hochholdinger, and Thomas Lubberstedt

This article is available at Iowa State University Digital Repository: https://lib.dr.iastate.edu/agron_pubs/455

Association analysis of single nucleotide polymorphisms in candidate genes with root traits in maize (*Zea mays* L.) seedlings

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

[1]Department of Agronomy, Agronomy Hall, Iowa State University, Ames, IA 50011, USA

[2] Department of Plant Production, Faculty of agriculture, Mu'tah University, P. O. Box 7, Karak, Jordan

[3]Science Institute, Autonomous University of Puebla (BUAP), Mexico

[4]INRES, Institute of Crop Science and Resource Conservation, Chair for Crop Functional Genomics, University of Bonn, 53113 Bonn, Germany

[†] These authors contributed equally to this work

*Corresponding author

Bharath Kumar: nbharath@iastate.edu

Adel H. Abdel-Ghani: abdelghani@mutah.edu.jo

Jordon Pace: jmpace1@iastate.edu

Jenaro Reyes-Matamoros: jenaro.reyes@correo.buap.mx

Frank Hochholdinger: hochhold@uni-bonn.de

Thomas Lübberstedt: thomasl@iastate.edu

Abstract

Background: Root growth and development is not only critical for nitrogen acquisition in plants, but also to anchor the plant in the soil. Several genes involved in maize root development have been isolated. Identification of SNPs associated with root traits would enable the selection of maize lines with better root architecture that might help to improve N uptake, and consequently plant growth particularly under N deficient conditions.

Results: In the present study, an association study (AS) panel consisting of 74 maize inbred lines was screened for seedling root traits in 6, 10, and 14-day-old seedlings. Allele re-sequencing of candidate root genes *Rtcl*, *Rth3*, *Rum1*, and *Rul1* was also carried out in the same AS panel lines. All four candidate genes displayed different levels of nucleotide diversity, haplotype diversity and linkage disequilibrium. Nucleotide diversity was highest in the *Rtcl* gene (π=0.021), intermediate in *Rum1* (π=0.011), lowest in *Rth3* (π=0.007) and *Rul1* (π=0.005) gene. When coding and non-coding regions within the genes were compared, nucleotide diversity varied across the genes. Gene based association analyses were carried out between individual polymorphisms in candidate genes, and root traits measured in 6, 10, and 14-day-old maize seedlings. Association analyses revealed several polymorphisms within the *Rtcl*, *Rth3*, *Rum1*, and *Rul1* genes associated with seedling root traits. These significantly associated SNPs also affected putative functional sequence motifs, mostly transcription factor binding sites, and major domains in the genes.

Conclusion: Several nucleotide polymorphisms in *Rtcl*, *Rth3*, *Rum1*, and *Rul1* were significantly ($P<0.05$) associated with seedling root traits in maize suggesting that all four tested genes are involved in the maize root development. Thus considerable allelic variation present in these root genes can be exploited for improving maize root characteristics. Target nucleotide

polymorphisms for functional marker development were identified which might find application in marker-based selection strategies in breeding programs.

Keywords

Maize, Root traits, Single nucleotide polymorphism, Nucleotide diversity, Linkage disequilibrium, Gene based association mapping

Background

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

Root architecture traits can be determined using different methods including vertical root pulling force (RPF) and hydroponic characterization [15-18]. Field methods are frequently technically demanding and costly. Due to the difficulty in obtaining reliable root trait data from the field, there are very few reports on morphological characterization of maize roots in the field. Using paper rolls as a hydroponic method to study root architecture has several advantages in comparison with RPF and other field techniques [7, 14, 18-19]. These include: (i) the ease to

score root traits as compared with vertical RPF, (ii) controlled environmental conditions, thus increasing repeatability of measurements, (iii) screening large numbers of lines in small space within a short period of time with an easy access to roots, and (iv) precise control of the concentration of mineral nutrients and water soluble compounds. However, the main disadvantages are the artificial screening conditions which might not properly represent field conditions.

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

The maize root system consists of different root types that are formed during different stages of plant development. The root system in maize can be divided into embryonic and post-embryonic roots [20]. The embryonic root system is composed of a single primary root and a variable number of seminal roots, while post-embryonic roots are shoot-borne roots including crown and brace roots. Shoot-borne roots formed at consecutive underground nodes are called crown roots, while the respective roots formed at consecutive above-ground nodes of the shoot are called brace roots. Lateral roots which emerge from all major root-types also belong to the post-embryonic root system. Mutants affected in various aspects of root formation have been identified in maize including *rtcs*, *rth1*, *rth3,* and *rum1* [19, 21-23]. *Rtcl* (*Rtcs-like*) is regarded as a paralog of *Rtcs* [22], and *Rul1* (*Rum1-like*) as a paralog of *Rum1*. That the primary root and its lateral roots alone are sufficient to form a fertile mature plant was demonstrated by the monogenic recessive mutant *rtcs*, which forms only a primary root and its lateral roots but no seminal or shoot-borne roots [24]. The mutant *rum1* is affected in lateral root formation, while

the mutants *rth1* [21] and *rth*3 [25] display reduced root hair elongation. *Rtcs* encodes a 244 amino acid (aa) Lateral Organ Boundaries (LOB) domain protein located on chromosome 1S. During evolution, *Rtcs* was duplicated. The *Rtcl* gene, which maps on chromosome 9, displays 72% aa sequence similarity with *Rtcs*. The *Rtcs* and *Rtcl* gene promoters share auxin responsive elements, and they are preferentially expressed in roots [22]. The *Rth1* gene encodes a SEC3 homologue [21]. **In yeast (*Saccharomyces cerevisiae*) and mammals**, *sec3* is part of the exocyst complex, which ropes together exocytotic vesicles prior to their fusion. The *Rth3* gene belongs to the COBRA-like gene family [25]. Members of this plant-specific glycosylphosphatidylinositol anchored protein coding gene family are involved in cell expansion and cell wall biosynthesis [26]. The *Rum1* gene located on chromosome 3 encodes a polypeptide of 269 aa which is a monocot specific AUX/IAA protein [23]. *Rul1* is a closely related Aux/IAA protein coding gene, and is localized on chromosome 8. *Rul1* encodes a polypeptide of 273 aa that displays 92% aa identity with *Rum1*.

Recombinant inbred lines have been used for mapping quantitative trait loci (QTL) to 10-30 cM regions [27-28]. However, QTL mapping is limited by, (i) the expense of generating such lines, (ii) their limited diversity, (iii) their separation from established processes in maize breeding, and (iv) the low number of informative recombinations [29]. In contrast, association mapping studies which are based on linkage disequilibrium (LD) allow identification of actual genes underlying these QTLs [30]. The success of gene based association studies depends on the candidate gene(s) chosen for a particular phenotypic trait. The first candidate gene-based association mapping study in maize associated individual *dwarf8* polymorphisms with flowering time [30], which has been followed by numerous subsequent studies in maize [31] and other cereal crops [32]. Gene-based association studies ultimately lead to the identification of quantitative trait polymorphisms

(QTPs) with causal genetic effects on agronomic traits, which can be converted into functional markers [33]. Breeding for a vigorous root system in maize may involve identification of superior alleles of candidate genes that affect nutrient and water use efficiency. Respective candidate gene-based studies enabled identification of alleles affecting various relevant quantitative agronomic traits in maize [30, 34-38].

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

Results

Phenotypic variation

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

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

Sequence alignment and haplotypes

The *Rtcl* sequence alignment of 69 maize lines spanned 828 bp with no alignment gaps, such as indel polymorphisms. The 828 bp amplified fragment included two exons, i.e., exon 1 (420 bp), and exon 2 (279 bp), respectively, separated by an intron (129 bp). In exons 1 and 2, 16 and 22 SNPs were identified, respectively, whereas 7 SNPs were identified in the intron region. Out of the 38 SNPs in the exon regions, 32 altered the amino-acid sequences; the other 8 were synonymous mutations (Table 1). In case of *Rth3*, 714 bp of the open reading frame (ORF) region of the gene was amplified from all 74 lines in the AS panel. Sequence alignment of *Rth3* amplicons resulted in the identification of 15 SNPs with no indel polymorphisms. Out of 15 SNPs, 12 were synonymous mutations, and only 3 SNPs altered the amino-acid sequence (Table 2). Intron 4 and exon 5 were partially amplified for *Rum1*: 225 out of 461 bp in intron 4 and 207 out of 315 bp in exon 5. Sequence alignment of *Rum1* amplicons from all 74 lines in the AS panel resulted in the identification of 12 SNPs. Out of 12 SNPs, 9 SNPs were from intron 4 and remaining 3 SNPs were from exon 5 region. Out of 3 SNPs identified in the exon region of *Rum1* gene, two were synonymous mutations and the remaining one SNP altered the amino acid sequence (Table 3). For *Rul1*, intron 5 and exon 6 were partially amplified from all 74 lines of the AS panel. Sequence alignment of 411 bp from *Rul1* consisting of 84 bp of intron 5 and 327 bp of exon 6 resulted in the identification of six SNPs, including five in exon 6. Four of those exon SNPs altered the amino-acid sequence (Table 4).

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

Nucleotide diversity in four genes

Nucleotide diversity (π) was determined for *Rtcl, Rth3, Rum1,* and *Rul1* coding and non-coding regions using the SNPs identified in respective amplicons from AS panel lines (Table 5). Overall, nucleotide diversity was π=0.021 in the entire region of *Rtcl*. Within *Rtcl*, nucleotide diversity was almost the same in both intron (π=0.022) and exon (π=0.021) regions. In *Rth3* which lacks an intron region, nucleotide diversity was higher for synonymous (π=0.026) than for non-synonymous mutations (π=0.0002). For the *Rum1*, nucleotide diversity was higher in the non-coding region (π=0.017) than in the coding region (π=0.005), and for *Rul1*, there was not much difference in the nucleotide diversity between non-coding (π=0.007) and coding region (π=0.004). When the entire amplified region was considered, nucleotide diversity was lower in *Rum1* (π=0.011), *Rth3* (π=0.007), and *Rul1* (π=0.005) compared to *Rtcl* (π=0.021). The nucleotide diversity based on θ, the neutral mutation parameter, was also calculated for all four amplicons in a sliding window of 100 bp using a step size of 10 bp (Fig. 1). Based on θ, within *Rtcl*, average nucleotide diversity was same in both intron and exon region. In case of the *Rum1* gene, nucleotide diversity seems to be higher in the intron region compared to the exon, but it was the same in both the exon and the intron region in the amplified region of the *Rul1* gene. Haplotype diversity (*Hd*) ranged from 0.873 in *Rtcl* to 0.624 in *Rul1*.

Tajima's *D* was positive and significant when considering the entire *Rtcl* region as well as both coding and non-coding regions. Conversely, in case of *Rth3*, *Rum1* and *Rul1*, Tajima's *D* was non-significant in all regions except in the *Rul1* non-coding region (Table 5). Complete analysis of LD decay in AS panel lines across *Rtcl*, *Rth3*, *Rum1* and *Rul1* genes has been reported elsewhere [54]. LD between all pairs of polymorphic sites from the sequenced region of the *Rtcl*,

Rum1, and *Rul1* genes decayed very rapidly ($r^2 < 0.2$), whereas LD persisted ($r^2 > 0.25$) over the length of the sequenced region in the *Rth3* gene.

Population structure and marker-trait associations

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

Rtcl

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

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

Rth3

13 polymorphisms in the *Rth3* exon region were associated with sTRL, whereas the SNP at 621 was the only polymorphism associated with sRDW (Table 7). Of these associated SNPs, a synonymous polymorphism at the site 621 was associated with both sTRL and sRDW. In case of fTRL and fRDW, seven and eight SNPs were associated, respectively. The synonymous SNPs at sites 180, 234, 438, 465, 492, 519, and 600 were significantly associated with both fTRL and fRDW. Moreover, these SNPs were also associated with sTRL. No SNP in *Rth3* was associated with tTRL. Four SNPs (389, 399, 436 and 600) significantly associated with TRL and RDW affected the binding sites for regulatory factors in the *Rth3* (Supp. Table 1). Since these SNPs were synonymous, they did not affect the COBRA domain in the *Rth3* gene.

Rum1

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

Rul1

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

Discussion

High levels of phenotypic, nucleotide, and haplotype diversity

We observed substantial quantitative variation for root traits TRL and RDW in 6, 10, and 14-day-old seedlings indicating a considerable amount of morphological differences among 74 maize inbred lines in the AS panel (Supp. Figs. 1 and 2). We identified maize lines with both under and well–developed root systems, which are attractive for identifying genomic regions controlling root traits [14].

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

intermediate frequencies and a scarcity of rare alleles. Considerable haplotype diversity was found for *Rtcl*, *Rth3*, *Rum1,* and *Rul1* (Table 5).

Polymorphisms associated with root traits

Several studies have shown the quantitative and qualitative importance of root traits in taking up nitrogen (N) from N-depleted soils [58-60]. Identification of the genetic regions associated with root traits would help not only to develop maize lines with a favorable root system, but also to understand the relationship between plant growth, plant productivity and root architecture. In our previous study, we identified significant positive correlations between seed root traits such as SRL and RDW with grain yield under two N levels [7]. Here, we used association mapping to dissect the role of SNPs in *Rtcl*, *Rth3*, *Rum1,* and *Rul1* for maize root development.

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

SNPs in *Rth3* were significantly associated with TRL and RDW in 6 and 14–day-old seedlings. Even though root hair elongation was not measured in this association study, our study suggests that *Rth3* affects other root characteristics in maize. Our findings are consistent with findings of Hochholdinger et al. [25], showing significant yield losses of the *rth3* mutant in replicated field trials. *Rth3* belongs to COBRA – like gene family specifically involved in cell expansion and cell wall biosynthesis [25-26]. The *rth3* mutant has been shown to affect root hair elongation and grain yield [25]. By this association mapping study we found that *Rth3* affects both TRL and RDW in maize seedlings. The significant association between SNPs in the *Rth3* gene with root length and biomass might be due to the role of root hairs in water and nutrient uptake. Previous studies have shown that plants lacking efficient uptake of water and nutrient have poor root characteristics [59, 65].

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

Molecular physiological basis of SNP– trait associations

Previous studies have shown the potential role of *Rth3* and *Rum1* genes in maize root development. Any impaired expression of these genes leads to defective root development. From

our gene based association study, we not only confirmed the role of *Rth3* and *Rum1* genes in maize root development, but we also found that the two paralogous genes *Rtcl* and *Rul1* are involved in the maize root formation. Thus, it is conceivable that polymorphisms in *Rtcl*, *Rth3*, *Rum1,* and *Rul1* affect maize root formation.

In the *Rtcl* gene, 13 non-synonymous and 4 synonymous SNPs were associated with TRL and RDW. Out of these associated SNPs, seven affected putative functional sequence motifs, mostly transcription factor binding sites. Moreover, out of these seven SNPs, three SNPs at sites 290, 317, and 320 also affected the LOB domain in the *Rtcl* gene. These SNPs seem to be critical not only for the formation of a proper LOB domain, which is required for root formation, but also for regulation of the *Rtcl* gene by affecting transcription factor binding sites. Similar results wherein the SNPs associated with traits affect transcription factor binding sites in the gene have been reported elsewhere [36, 43]. In our previous association mapping study involving SNPs from the *Rtcl* gene and seedling root traits measured under contrasting nitrogen levels, these three SNPs were consistently associated with seedling root traits. The SNP at site 317 in *Rtcl* gene was associated with both RDW and TRL under high and low N conditions, whereas the SNP at site 320 was associated with RDW under both N conditions. In case of the SNP at site 290, associations were observed with RDW and TRL under high N. These consistent associations suggest the potential role of these SNPs in the *Rtcl* gene in maize root development. LD is very low between SNPs at sites 290-320 (r^2 = 0.0255) and 317-320 (r^2 = 0.0903), whereas it was moderate between 290-317 (r^2 = 0.2827). Low to moderate LD between these significant SNPs suggests that these individual SNPs might be true causative polymorphisms, and can be of potential use in deriving markers to select root traits.

For the *Rth3* gene, 13 polymorphisms were found to be significantly associated with TRL and RDW. Of these 13 SNPs, four SNPs (site 393, 399, 438, and 600) in the exon region significantly affected the binding sites for regulatory factors in *Rth3*, but none of these SNPs affected the COBRA domain within the gene, as they were synonymous mutations. When the LD was estimated between these four SNPs, low LD was detected between SNPs at the sites 393-438 (r^2 =0.0445), 393-600 (r^2 =0.0323), 399-438 (r^2 =0.0445) and 399-600 (r^2 =0.0323). However, there was a high LD between sites at positions 393-399 and 438-600. This suggests, that individual SNPs at 393 (or 399), 438 (or 600) are the true causative polymorphisms, and can potentially be used to derive markers to select root traits. In our previous association mapping study involving SNPs in the *Rth3* gene and grain yield, the SNP at site 600 was associated with grain yield under high N suggesting that this SNP might potentially be used along with other SNPs to select for grain yield. In case of the *Rth3* gene, full-length re-sequencing of this candidate gene would greatly increase the number of unlinked polymorphisms to be tested for associations due to the extent of LD over a long distance.

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

In the present study, genes and their paralogues have been tested for association with roots traits. From our results, it seems that *Rtcl*, *Rth3*, *Rum1,* and *Rul1* can be considered as candidate genes to develop functional markers for root traits especially the significant SNPs in these genes with large effect on the trait (Supp. Table 2). Functional markers are DNA markers derived from polymorphic sites within genes, causally involved in phenotypic trait variation [33]. One of our future objectives is to develop functional markers for seedling root traits, and to validate them in independent association mapping populations.

Methods

Plant materials

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

Experimental design and phenotyping

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

DNA extraction, amplification, and sequencing

Four candidate genes were chosen based on their role in root development to identify SNPs for association study analyses. SNPs from these candidate genes were tested for possible

associations with TRL and RDW. Candidate genes chosen for our association study were: *Rtcs*, *Rtcn*, *Rtcl*, *Rth3*, *Rum1,* and *Rul1*. Gene specific primers were designed to amplify the entire sequence of *Rtcl,* and parts of *Rth3*, *Rum1,* and *Rul1* genes using the software program Primer 3.0 (http://frodo.wi.mit.edu/primer3/) (Table 3). In case of *Rtcs* and *Rtcn*, even after several attempts, amplicons from all 74 lines were not obtained. This might be due to the extensive nucleotide diversity at these candidate genes which prevents the binding of designed primers. Polymerase chain reaction (PCR) was performed using the designed primers for each gene separately in 50 μl volumes under the following conditions: 50 ng template DNA, 250 nM of each primer, 250 nM dNTPs, 2 U Taq polymerase and 250 μM $MgCl_2$. Reactions were performed for each primer pair using the following PCR program in a thermocycler (MJ research, California): an initial 94 °C denaturation step for 2 min followed by 35 cycles of 94 °C for 30 sec (denaturation step), 57.5 °C for 30 sec (annealing step), and 72 °C for 90 sec (elongation step). The final extension step was followed by 72 °C incubation for 10 min. Amplified DNA fragments were resolved by gel electrophoresis (Biorad, California) using 1% agarose gels in Tris-EDTA (TE) buffer. Agarose gels were stained with 0.5 µg of ethidium bromide per ml. The running time was 90 min at 120 mV. Finally, gels were visualized and photographed by a UV illuminator system (Alphainnotech, California). For each gel, the first lane was specified for a 100 bp DNA ladder (Promega, Wisconsin), the second lane and the third lane were specified for positive and negative controls. Amplified fragments of *Rth3, Rum1,* and *Rul1* genes were obtained for all 74 inbred lines in the AS panel, whereas for the *Rtcl* gene, amplicons were obtained from 69 lines. For sequencing, 10 µl of the amplified fragments were first purified by using 2 units of shrimp alkaline phosphatase and 2 units of exonuclease I at 37 °C for 1 h, followed by 72 °C for 15 min to deactivate the enzymes. Amplified gene products

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

Phenotypic data analyses

The following linear mixed model was used to estimate variance components: $y_{ijk} = \mu + E_i + B_{j(i)} + G_k + EG_{ik} + BG_{j(i)}k$, where y_{ijk} represents the observation from the ijk^{th} experimental unit, μ is the overall mean, E_i is the effect of ith independent experiment, $B_{j(i)}$ is the effect of jth block nested in i^{th} experiment, G_k is the effect of kth line, EG_{ik} is the interaction effect of the ith experiment with k^{th} genotype, $BG_{j(i)k}$ is the interaction effect of jth block nested in i^{th} experiment with k^{th} genotype. Heritability (h^2) on an entry mean basis was estimated as the ratio of genotypic to phenotypic variance according to Hallauer and Miranda [39]. Furthermore,

experiment was considered as a fixed factor, whereas blocks and genotypes were regarded as random factors. Best linear unbiased estimates (BLUEs) were determined for maize lines for each trait. SAS 9.1 (SAS Institute, 1996) software packages were used for all calculations.

Analysis of sequence data

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

Population structure and association analysis

All 74 lines in the AS panel were genotyped with 101 SNP markers distributed evenly across 10 maize chromosomes [46] to asses and control the effect of population structure. The software package *Structure 2.0* [47] was used to estimate population structure (Q) within the AS panel using SNP data. In *Structure 2.0*, a burn-in length of 50.000 followed by 50.000 iterations for

each of the clusters (K) varying from 1 to 20 (each K was run 20 times) were used to produce a Q matrix estimating membership coefficients for each line in each subpopulation. The Admixture model was applied with independent allele frequencies. An ad hoc (ΔK) statistic [48] based on the second order rate change of P (X|K) was used to identify the most probable value of K.

Loiselle kinship coefficients between lines (a K matrix) were estimated by the TASSEL program [49] based on the 101 SNP markers. Both Q matrix and a K matrix were used in the association analysis to control the spurious associations due to population structure and relatedness, respectively [50]. Association analysis between SNPs and root traits was carried out using a mixed linear model (MLM) implemented in the program *TASSEL 2.10* software [49]. The MLM accounts for overall population structure (Q) and for finer scale relative kinship (K). The statistical model used in mixed linear model (Q+K) can be described in Henderson's notation [51] as:

$y= X\beta+Zu+e$, where y is the vector of observations; β is an unknown vector containing fixed effects including genetic marker and population structure (Q); u is an unknown vector of random additive genetic effects from multiple background QTL for individuals or lines; X and Z are the known design matrices; and e is the unobserved vector of random residuals.

TRL and RDW were measured in 6 (sTRL, sRDW), 10 (tTRL, tRDW), and 14 (fTRL, fRDW) day old seedlings, and used as root traits in our association study. False discovery rate was set at 0.05 [52] to control for multiple testing of SNP markers. Motifs in the *Rtcl*, *Rth3*, *Rum1,* and *Rul1* genes were searched using a PLACE (Plant cis-acting regulatory DNA elements) database [53] to determine, if any of the significantly associated SNPs might alter motif sequences in the candidate genes.

Additional material

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

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

Supplementary Table 1: Significantly associated SNP's affecting motifs in the candidate genes.

Supplementary Table 2: Association analysis results

Abbreviations

Rtcs: rootless concerning crown and seminal roots; *Rtcl*: *Rtcs*-like protein; *Rth3:* roothairless 3; *Rum1*: rootless with undetectable meristems 1; *Rul1*: *Rum1*-like 1; SNP: single-nucleotide polymorphism; LD: linkage disequilibrium; TRL; total root length; RDW: root dry weight.

Author's contribution

BK prepared the manuscript with the help of AHA. BK designed the experiments, and with the help of FH setup the phenotyping experiments. BK and JR were involved in the root measurements. BK and AHA with a help of JP carried out allele sequencing. BK performed data analysis. AHA and FH reviewed the manuscript. TL coordinated the project and together with BK prepared the manuscript. All authors read and approved the final manuscript.

Author's information

[1]Department of Agronomy, Agronomy Hall, Iowa State University, Ames, IA 50011, USA

[2] Department of Plant Production, Faculty of agriculture, Mu'tah University, P. O. Box 7, Karak, Jordan

[3]Science Institute, Autonomous University of Puebla (BUAP), Mexico

[4]INRES, Institute of Crop Science and Resource Conservation, Chair for Crop Functional Genomics, University of Bonn, 53113 Bonn, Germany

Acknowledgements

This study was prepared while Dr. Adel Abdel-Ghani was a visiting Fulbright Postdoctoral Fellow and during the sabbatical leave granted to Dr. Adel Abdel-Ghani from Mu'tah University, Jordan during the academic year 2011-2012 at Iowa State University (ISU), Ames, USA. Dr. Jenaro Reyes-Matamoros contributed in this study while he was on a sabbatical leave at ISU. Authors are very thankful to undergraduate student helpers Guan Yi and Leigh Lihs for their technical assistance during this work. Authors would also like to thank USDA's National Institute of Food and Agriculture (project number: IOW05180) for funding this work. Root work in F.H.s laboratory is supported by the DFG (Deutsche Forschungsgemeinschaft). Financial assistance from RF Baker Center for Plant Breeding is highly appreciated.

References

1. Tuberosa R, Salvi S: **From QTLs to genes controlling root traits in maize. In: Spiertz JHJ, Struik PC, Van Laar HH (eds.) Scale and complexity in plant systems research: gene–plant–crop relations**. *Springer* 2007, pp 13–22.
2. Mooney HA, Bloom AJ, Chapin III FS: **Resource limitation in plants-an economic analogy**. *Annu Rev of Ecol Syst* 1985, **16**:363-392.
3. Robinson D: **Compensatory changes in the partitioning of dry matter in relation to nitrogen uptake and optimal variations of growth**. *Ann Bot-London* 1986, **58**:841-848.
4. Hilbert DW: **Optimization of plant root: shoot ratios and internal nitrogen concentration.** *Ann Bot-London* 1990, **66**:91-99.

5. Reynolds HL, D'Antonio C: **The ecological significance of plasticity in root weight ratio in response to nitrogen: Opinion**. *Plant Soil* 1996, **185**:75-97.
6. Bonifas KD, Walters DT, Cassman KG, Lindquist JL: **Nitrogen supply affects root: shoot ratio in corn and velvetleaf (*Abutilon theophrasti*)**. *Weed Sci* 2005, **53**:670-675.
7. Abdel-Ghani AH, Kumar B, Reyes-Matamoros J, Gonzalez-Portilla PJ, Jansen C, San Martin JP, Lee M, Lübberstedt T**: Genotypic variation and relationships between seedling and adult plant traits in maize (*Zea mays* L.) inbred lines grown under contrasting nitrogen levels**. *Euphytica* 2013, **189**: 123-133.
8. Nye PH, Tinker PB: **Solute Movement in the Soil-Root System**. *University of California press, Berkeley*, 1977.
9. Sattelmacher B, Klotz F, Marschner H: **Influence of the nitrogen level on root growth and morphology of two potato varieties differing in nitrogen acquisition**. *Plant Soil* 1990, **123**:131-137.
10. Hurd EA, Townley-Smith TF, Patterson LA, Owen CH: **Techniques used in producing Wascana wheat**. *Can J Plant Sci* 1972, **52**:689–691.
11. Barber SA, MacKay AD: **Root growth and phosphorus and potassium uptake by two corn genotypes in the field**. *Fert Res* 1986, **10**:217–230.
12. Gregory PJ: **Root growth and activity. In Peterson GA (ed) Physiology and determination of crop yield**. ASA, CSSA, and SSSA, Madison, WI, 1994, pp 65–93.
13. Marschner H: **Role of root growth, arbuscular mycorrhiza, and root exudates for the efficiency in nutrient acquisition**. *Field Crops Res* 1988, **56**:203–207.

14. Kumar B, Abdel-Ghani AH, Reyes-Matamoros J, Hochholdinger F and Lübberstedt T: **Genotypic variation for root architecture traits in seedlings of maize (Zea mays L.) inbred lines**. *Plant Breeding* 2012, **131**:465-478.
15. Andrew RH, Solanki SS: **Comparative root morphology for inbred lines of corn as related to performance**. *Agron J* 1966, **58**:415-418.
16. Zuber MS: **Evaluation of corn root system under various environments.23rd Corn and Sorghum Conf.** 1968, pp 1-19.
17. Nass HG, Zuber MS: **Correlation of corn (*Zea mays* L.) roots to early in development to mature root development**. *Crop Sci* 1971, **11**:655-658.
18. Tuberosa R, Salvi S, Sanguinetti MC, Maccaferi M, Giuliani S, Landi P: **Searching for quantitative trait loci controlling root traits in maize: a critical appraisal**. *Plant Soil* 2003, **255**:35-54.
19. Woll K, Borsuk L, Stransky H, Nettleton D, Schnable PS, Hochholdinger F: **Isolation, characterization and pericycle specific transcriptome analyses of the novel maize (*Zea mays* L.) lateral and seminal root initiation mutant rum1**. *Plant Physiol* 2005, **139**:1255-1267.
20. Abbe EC, Stein OL: **The origin of the shoot apex in maize: embryogeny**. *Am J Bot* 1954, 41:285-293.
21. Wen TJ, Hochholdinger F, Sauer M, Bruce W, Schnable PS: **The *roothairless1* gene of maize (*Zea mays*) encodes a homolog of *sec3*, which is involved in polar exocytosis**. *Plant Physiol* 2005, **138**:1637-1643.
22. Taramino G, Sauer M, Stauffer J, Multani D, Niu X, Sakai H, Hochholdinger F: **The *rtcs* gene in maize (*Zea mays* L.) encodes a lob domain protein that is required**

for postembryonic shoot-borne and embryonic seminal root initiation. *Plant J* 2007, **50**:649-659.

23. Von Behrens I, Komatsu M, Zhang Y, Berendzen KW, Niu X, Sakai H, Taramino G, Hochholdinger F: **Rootless with undetectable meristem1 encodes a monocot-specific AUX/IAA protein that controls embryonic seminal and postembryonic lateral root initiation in maize**. *Plant J* 2011, **66**: 341-353.
24. Hertz W, Hochholdinger F, Schwall F, Feix G: **Isolation and characterization of *rtcs*, a mutant deficient in the formation of nodal roots**. *Plant Journal* 1996, **10**:845-857.
25. Hochholdinger F, Wen TJ, Zimmermann R, Chimot-Marolle P, da Costa e Silva O, Bruce W, Lamkey KR, Wienand U, Schnable PS: **The maize (*Zea mays* L.) *roothairless3* gene encodes a putative GPI-anchored, monocot-specific, COBRA-like protein that significantly affects grain yield**. *Plant J* 2008, **54**:888.
26. Brady SM, Song S, Dhugga KS, Rafalski JA, Benfey PN: **Combining Expression and Comparative Evolutionary Analysis. The COBRA Gene Family**. *Plant Physiol* 2007, **143**:172-187.
27. Alpert KB, Tanksley SD: **High-resolution mapping and isolation of a yeast artificial chromosome contig containing *fw2.2*: a major fruit weight quantitative trait locus in tomato**. *Proc Natl Acad Sci USA* 1996, **93**:15503–15507.
28. Stuber CW, Polacco M, Senior ML: **Synergy of empirical breeding, marker-assisted selection, and genomics to increase crop yield potential**. *Crop Sci* 1999, **39**:1571-1583.

29. Jannink J, Lorenz AJ, Iwata H: **Genomic selection in plant breeding: from theory to practice**. *Brief Funct Genomics* 2010, **9**:166-177.
30. Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES: **Dwarf8 polymorphisms associate with variation in flowering time**. *Nat Genet* 2001, **28**:286-289.
31. Yu J, Buckler ES: **Genetic association mapping and genome organization of maize**. *Curr Opin Biotech* 2006, **17**:155–160.
32. Gupta PK, Rustgi S, Kulwal, PL: **Linkage disequilibrium and association studies in higher plants: Present status and future prospects**. *Plant Mol Biol* 2005, **57**:461-485.
33. Andersen JR, Lübberstedt T: **Functional markers in plants**. *Trends Plant Sci* 2003, 8:554 –560.
34. Palaisa KA, Morgante M, Williams M, Rafalski A: **Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci**. *Plant Cell* 2003, **15**:1795-1806.
35. Andersen JR, Zein I, Wenzel G, Krützfeldt B, Eder J, Ouzunova M, Lübberstedt T: **High levels of linkage disequilibrium and associations with forage quality at a phenylalanine ammonia-lyase locus in European maize (Zea mays L.) inbreds**. *Theor Appl Genet* 2007, **114**:307-19.
36. Brenner A, Zein I, Chen Y, Andersen JR, Wenzel G, Ouzunova M, Eder J, Darnhofer B, Frei U, Barrière Y, Lübberstedt T: **Polymorphisms in O-methyltransferase genes are associated with stover cell wall digestibility in European maize (*Zea mays* L.)**. *BMC Plant Biol* 2010, **10**: 27.

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

38. Setter TL, Yan J, Warburton M, Ribaut JM, Xu Y, Sawkins M, Buckler ES, Zhang Z, Gore MA: **Genetic association mapping identifies single nucleotide polymorphisms in genes that affect abscisic acid levels in maize floral tissues during drought**. *J Exp Bot* 2011, **62**: 701–716.

39. Hallauer AR, Miranda JB: **Quantitative genetics in maize breeding** 1981, 2nd edition, *Iowa State University Press*, Ames, Iowa, pp. 468.

40. Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R: **DnaSP DNA polymorphism analyses by the coalescent and other methods**. *Bioinformatics* 2003, **19**:2496-2497.

41. Nei M: **Molecular Evolutionary Genetics**. *Columbia University Press* 1987, New York, NY.

42. Watterson GA: **On the number of segregating sites in genetical models without recombination**. *Theor Popul Biol* 1975, **7**:256-276.

43. Li Y, Haseneyer G, Schön C, Ankerst D, Korzun V, Wilde P, Bauer E: **High levels of nucleotide diversity and fast decline of linkage disequilibrium in rye (Secale cereale L.) genes involved in frost response.** *BMC Plant Biology* 2011, **11**:6.

44. Tajima F: **Evolutionary relationship of DNA sequences in finite populations**. *Genet* 1983, **105**:437-460.

45. Tajima F: **DNA Polymorphism in a Subdivided Population: The Expected Number of Segregating Sites in the Two-Subpopulation Model.** *Genet* 1989, **123**:229-240.
46. Barbazuk WB, Emrich SE, Chen HD, Li L, Schnable PS: **SNP discovery via 454 transcriptome sequencing**. *Plant J* 2007, **51**:910-918.
47. Pritchard JK, Stephens M, Donnelly P: **Inference of Population Structure Using Multilocus Genotype Data**. *Genet* 2000, **155**:945-959.
48. Evano G, Regnaut S, Goudet J: **Detecting the number of clusters of individuals using the software structure: A simulation study**. *Mol Ecol* 2005 **14**:2611-2620.
49. 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**:2633-2635.
50. Zhang Z, Ersoz E, Lai C, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu J, Arnett DK, Ordovas JM, Buckler ES: **Mixed linear model approach adapted for genome wide association studies**. *Nat Genet* 2010, **42**: 355-60.
51. Henderson CR: **Best linear unbiased estimation and prediction under a selection model**. *Biometrics* 1975, **31**: 423–447.
52. Storey JD, Taylor JE, Siegmund D: **Strong control, conservative point estimation, and simultaneous conservative consistency of false discovery rates: A unified approach**. *J Roy Stat Soc B* 2004, **66**:187-205.
53. Higo K, Ugawa Y, Iwamoto M, Korenaga T: **Plant cis-acting regulatory DNA elements (PLACE) database**. *Nucleic Acids Research* 1999, **27**:297-300.

54. Abdel-Ghani AH, Kumar B, Jansen C, Gonzalez-Portilla PJ, Reyes-Matamoros J, San Martin JP, Lee M, Lübberstedt T: **Association analysis of genes involved in maize (Zea mays L.) root development with seedling and agronomic traits under contrasting nitrogen levels (unpublished)**.

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

56. Zein I, Wenzel G, Andersen JR, Lübberstedt T: **Low Level of Linkage Disequilibrium at the COMT (Caffeic Acid O-methyl Transferase) Locus in European Maize (Zea mays L.)**. *Genetic Resources and Crop Evolution* 2007, **54**:139-148.

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

58. Guingo E, Hébert Y, Charcosset A: **Genetic analysis of root traits in maize**. *Agronomie* 1998, **18**: 225–235.

59. Kamara AY, Kling JG, Ajala SO, Menkir A: **The relationship between vertical root-pulling resistance and nitrogen uptake and utilization in maize breeding lines**. *Maydica* 2002, **47**:135-140.

60. Coque M, Gallais A: **Genomic regions involved in response to grain yield selection at high and low nitrogen fertilization in maize**. *Theoretical and Applied Genetics* 2006, **112**: 1205–1220.

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

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

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

64. Inukai Y, Sakamoto T, Ueguchi-Tanaka M, Shibata Y, Gomi K, Umemura L, Hasegawa Y, Ashikari M, Kitano H, Matsuokaa M: **Crown rootless1, Which Is Essential for Crown Root Formation in Rice, Is a Target of an AUXIN RESPONSE FACTOR in Auxin Signaling**. *Plant Cell* 2005 **17**:1387-1396.

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

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

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

SNP position	Exon 1																Intron 1							Exon 2																					
	33	94	204	232	290	296	298	307	317	320	324	332	357	373	374	413	468	473	481	510	531	543	547	554	576	597	601	607	616	632	648	694	695	703	704	708	711	716	720	736	737	759	761	792	824
	s	a	s	a	a	a	a	a	a	a	s	a	s	a	a	a								s	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	s	a	a
Haplotypes																																													
Hap_1	G	C	C	C	T	A	A	T	C	C	G	C	A	G	G	C	A	A	G	A	G	A	T	C	C	G	G	N	C	G	A	G	G	C	G	A	A	C	T	A	T	C	T	T	A
Hap_2	G	C	C	C	T	A	A	T	C	C	C	C	A	G	A	C	C	G	A	T	G	A	T	C	C	G	G	N	C	G	A	G	G	C	G	A	A	C	A	C	T	C	T	G	A
Hap_3	G	C	C	C	T	A	A	T	T	C	C	A	A	G	A	C	A	A	G	A	G	A	T	C	C	G	A	N	T	G	T	G	T	C	G	T	A	A	A	N	A	C	C	G	C
Hap_4	A	G	A	C	T	A	A	T	C	T	G	A	A	G	G	C	A	A	G	A	G	A	T	C	C	G	G	N	C	G	A	G	G	C	G	A	A	C	T	A	T	C	T	T	A
Hap_5	A	G	C	C	A	G	T	T	T	C	T	C	T	A	N	A	T	G	G	T	C	G	G	T	C	T	G	N	C	A	A	G	G	G	G	A	A	C	A	C	T	C	T	A	C
Hap_6	G	C	C	C	A	G	T	T	T	C	T	C	T	A	N	A	T	G	G	A	G	A	T	C	C	G	G	N	C	G	A	G	G	C	G	A	A	C	A	C	T	C	T	G	A
Hap_7	G	C	C	C	T	A	A	T	C	C	C	C	A	G	A	C	A	A	A	T	G	A	T	T	C	G	G	N	C	G	G	G	G	C	G	A	A	C	A	N	A	G	T	T	C
Hap_8	G	C	C	A	T	A	A	T	T	C	C	C	T	A	N	C	A	A	A	T	G	A	T	T	C	G	G	N	C	G	G	G	G	C	G	A	A	C	A	N	A	G	T	T	C
Hap_9	G	C	C	C	T	A	A	T	C	C	C	C	A	G	A	C	N	A	A	T	G	A	T	T	C	G	G	N	C	G	G	G	G	C	G	A	A	C	A	N	A	G	T	T	C
Hap_10	G	C	A	C	T	A	A	T	C	T	G	A	A	G	G	C	A	A	G	T	G	A	T	C	C	G	G	N	C	G	G	A	G	N	A	T	T	C	A	N	A	C	T	T	C
Hap_11	G	C	C	C	T	A	A	T	C	T	G	A	G	G	G	C	A	A	G	T	G	A	T	C	C	G	G	N	C	G	G	A	G	N	A	T	T	C	A	N	A	C	T	T	C
Hap_12	G	C	A	A	T	A	A	A	C	C	C	C	A	G	G	A	T	G	G	T	G	A	G	T	T	G	G	G	C	G	A	A	G	N	A	T	T	C	A	N	A	C	T	T	C
Hap_13	G	C	C	C	T	A	A	T	T	C	C	A	A	G	A	C	A	A	G	A	G	A	T	C	C	G	A	C	T	G	T	G	T	C	G	T	A	A	A	N	A	C	C	G	C
Hap_14	G	C	C	C	T	A	A	T	T	C	C	A	A	G	A	C	A	A	G	A	G	A	T	C	C	G	A	N	T	G	T	G	T	C	G	T	A	N	A	N	A	C	C	G	C
Hap_15	A	G	C	C	A	G	T	T	T	C	T	C	T	A	N	A	T	G	G	T	G	G	G	T	C	T	G	N	C	A	A	G	G	G	G	A	A	C	A	C	T	C	T	A	C
Hap_16	G	G	C	C	T	A	A	T	T	C	C	A	A	G	A	C	A	A	G	T	G	G	G	T	C	T	G	N	C	A	A	G	G	G	G	A	A	C	A	C	T	C	T	A	C

2

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

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

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

SNP Position																					
								Exon													
	59a	163a	479a	180s	234s	351s	393s	399s	417s	438s	465s	492s	519s	600s	621s						
Haplotypes																sTRL	tTRL	fTRL	sRDW	tRDW	fRDW
Hap_1	C	G	G	C	G	G	G	C	G	C	A	G	G	T	G	39.9	128.8	215.6	14.74	34.13	50.32
Hap_2	C	A	G	C	G	G	G	G	A	C	A	G	G	T	G	42.7	144.9	235.6	15.96	32.69	52.23
Hap_3	N	N	G	C	G	G	G	G	A	C	A	G	G	T	G	40.9	134.8	278.9	16.51	36.29	62.89
Hap_4	C	G	T	T	G	G	G	C	G	A	A	T	A	C	A	34.1	130.2	204.0	11.26	28.21	42.32
Hap_5	C	G	G	C	T	A	A	C	A	C	A	G	G	T	A	32.1	143.8	250.4	11.88	29.38	52.05
Hap_6	C	G	G	C	G	G	G	G	A	C	A	G	G	T	G	35.9	141.3	224.9	12.89	29.67	48.22
Hap_7	T	G	G	C	G	G	G	G	A	C	A	G	G	T	G	45.7	150.5	230.0	16.69	33.23	50.20
Hap_8	N	G	G	C	G	G	G	G	A	C	A	G	G	T	G	20.1	149.1	194.2	12.41	29.99	40.60
Hap_9	C	A	G	C	G	G	G	G	A	C	G	G	G	T	G	46.3	143.7	247.0	14.64	33.01	58.76
															Maximum	46.3	150.5	278.9	16.69	36.29	62.89
															Minimum	20.1	128.8	194.2	11.26	28.21	40.60
															Range	26.2	21.7	84.7	5.43	8.08	22.29

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

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

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

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

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

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

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

1 **Table 5: Summary of alignment length, number of genotypes per alignment, polymorphisms and nucleotide diversity in the *Rtcl*, *Rth3*, *Rum1* and *Rul1*
2 genes in maize.**
3

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

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

5

6

7

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

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

TRL=Total Root Length; RDW = Root Dry Weight

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

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

TRL=Total Root Length; RDW = Root Dry Weight

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

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

TRL=Total Root Length; RDW = Root Dry Weight

1

2 **Table 9: Polymorphic sites of *Rul1* gene associated with the root traits (Total root length, Root dry weight) at**
3 **different growth stages identified by MLM analysis.**
4

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

5 RDW = Root Dry Weight

6

7

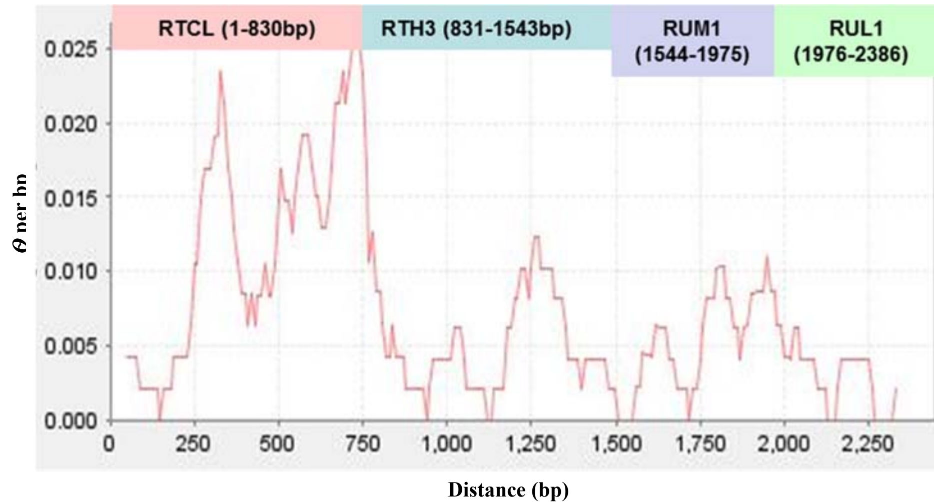


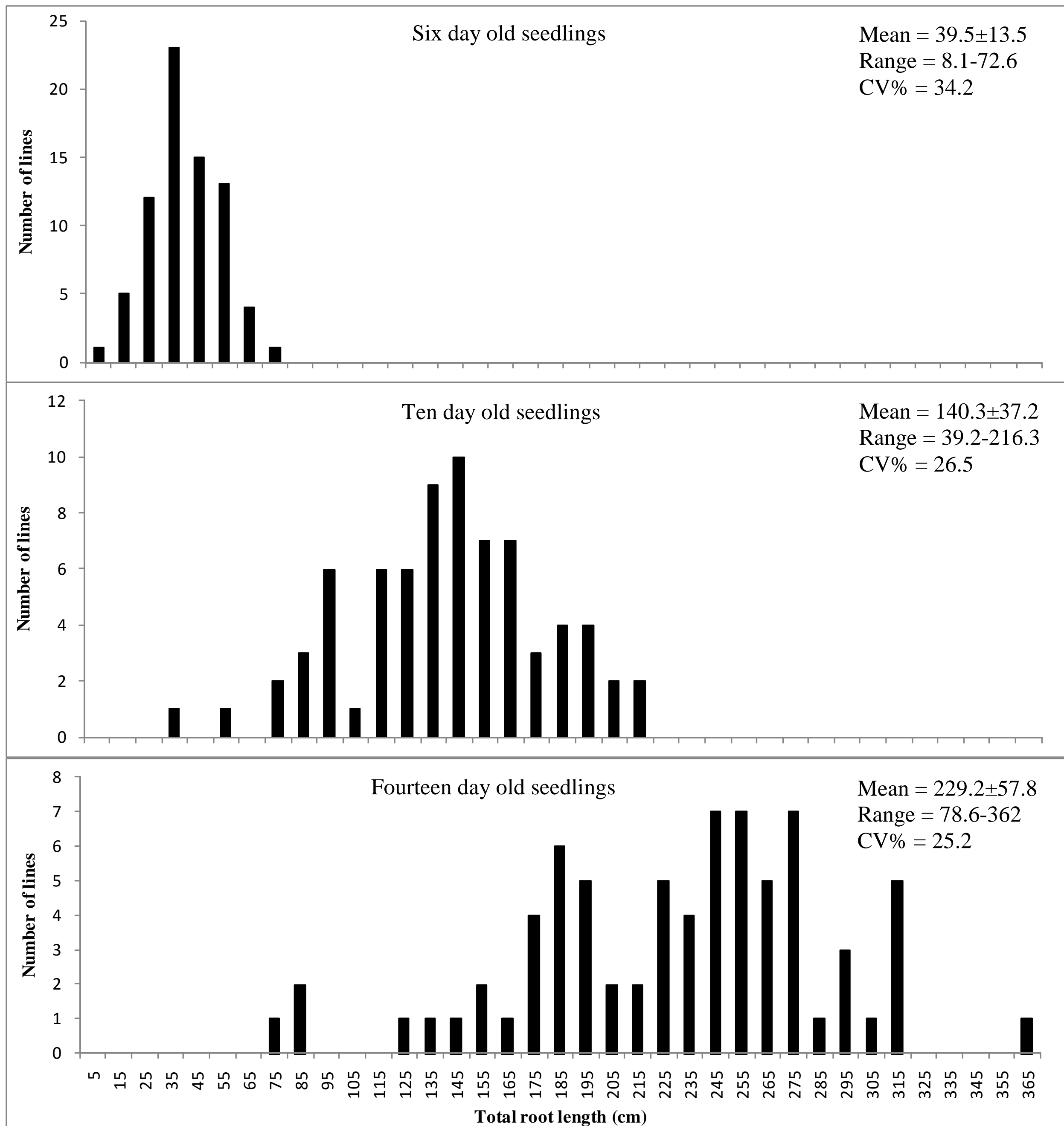
Figure 1: Nucleotide diversity values (x-axis) in sliding windows (size = 10, length = 100) at the *Rtcl, Rth3, Rum1* and *Rul1* gene locus for all lines. Nucleotide diversity was calculated based on the θ, neutral mutation parameter derived from the total number of segregating sites. 1-420bp: Rtcl Exon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

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

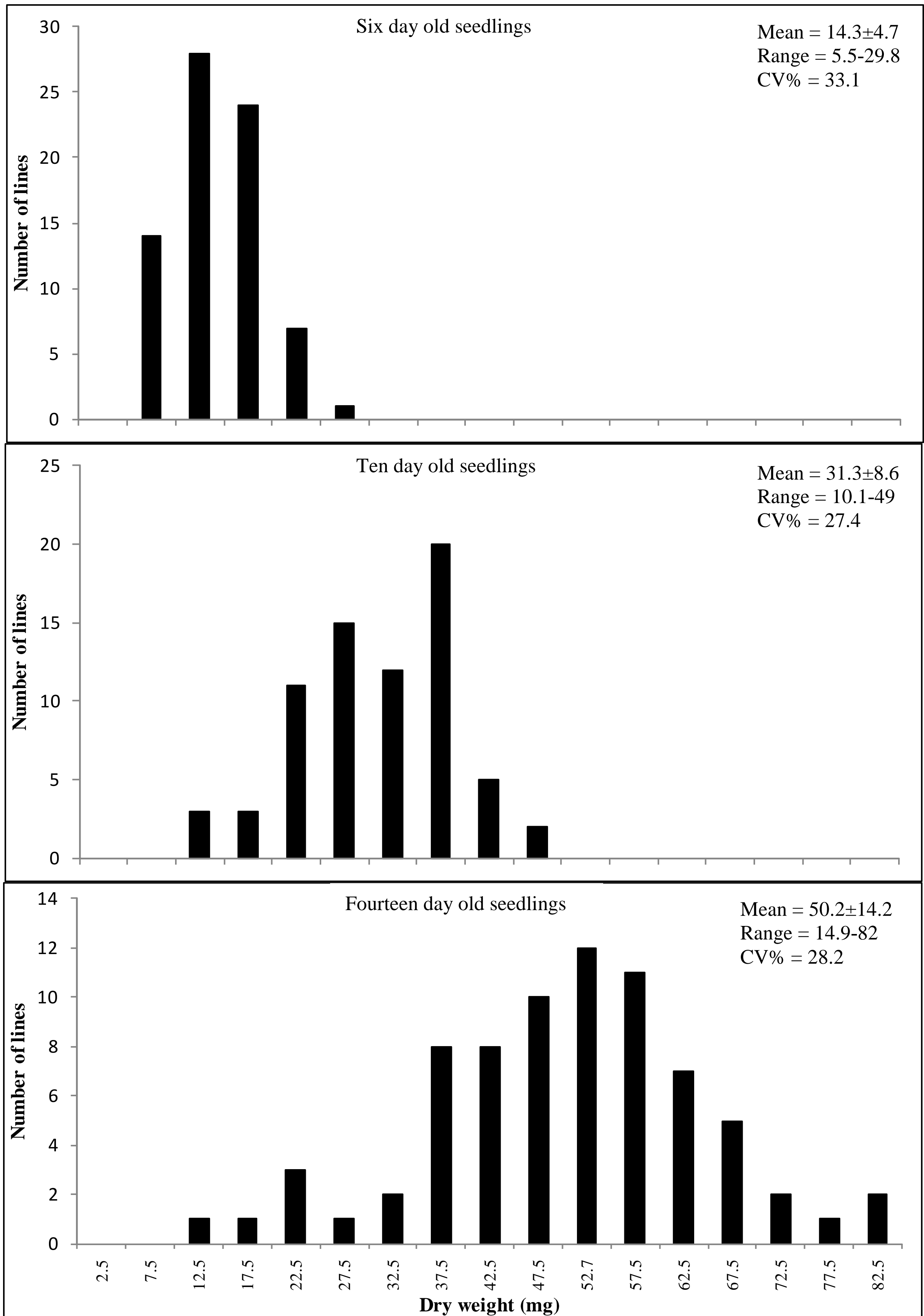
Supplementary figure 2: Distribution of trait dry weight in maize inbred lines, values in x-axis represents mid-point of class-interval.

Supplementary Table 1: Significantly associated SNP's affecting motifs in the candidate genes.

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



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



Supplementary Table 1: Significantly associated SNP's affecting motifs in the candidate genes.

Gene	Motif Location	Motif Name	Motif Sequence	SNP site	Nucleotide change	Motif Description
Rtcl	290	TBOXATGAPB	ACTTTG	290	T→A	Tbox found in the Arabidopsis thaliana GAPB gene promoter. Mutations in the "Tbox" resulted in reductions of light-activated gene transcription; GAPB encodes the B subunit of chloroplast glyceraldehyde-3-phosphate dehydrogenase (GADPH) of Arabidopsis thaliana.
Rtcl	317	AGCBOXNPGLB	AGCCGCC	317 320	C→T T→C	AGC box repeated twice in a 61 bp enhancer element in tobacco (N.p.) class I beta-1,3-glucanase (GLB) gene. Binding sequence of Arabidopsis AtERFs;AtERF1,2 and 5 functioned as activators of GCC box-dependent transcription; AtERF3 and 4 acted as repressors; AtERF proteins are stress signal-response factors; EREBP2 binding site; Conserved in most PR-protein genes; Rice MAPK (BWMK1) phosphorylates OS EREBP1, which enhance DNA-binding activity of the factor to the GCC box.
Rtcl	467	WRKY71OS	TGAC	468	A→T	A core of TGAC-containing W-box" of, e.g., Amy32b promoter; Binding site of rice WRKY71, a transcriptional repressor of the gibberellin signaling pathway; Parsley WRKY proteins bind specifically to TGAC-containing W box elements within the Pathogenesis-Related Class10 (PR-10) genes.
Rtcl	468	GTGANTG10	GTGA	468	A→T	GTGA motif" found in the promoter of the tobacco (N.t.) late pollen gene g10 which shows homology to pectate lyase and is the putative homologue of the tomato gene lat56.
Rtcl	507	ACGTATERD1	ACGT	510	T→A	ACGT sequence (from -155 to -152) required for etiolation-induced expression of erd1 (early responsive to dehydration) in Arabidopsis.
Rtcl	596	LTRECOREATCOR15	CCGAC	597	G→T	Core of low temperature responsive element (LTRE) of cor15a gene in Arabidopsis ; A portion of repeat-C (C-repeat),TGGCCGAC, which is repeated twice in cor15a promoter; ABA responsiveness; Involved in cold induction of BN115 gene from winter Brassica napus; Light signaling mediated by phytochrome is necessary for cold- or drought- induced gene expression through the C/DRE in Arabidopsis.
Rtcl	795	BIHD1OS	TGTCA	799	T→G	Binding site of OsBIHD1, a rice BELL homeodomain transcription factor;

Rth3	389	GCCCORE	GCCGCC	393	G→A	Core of GCC-box found in many pathogen-responsive genes such as PDF1.2, Thi2.1, and PR4; Has been shown to function as ethylene-responsive element; Appears to play important roles in regulating jasmonate-responsive gene expression; Tomato Pti4 (ERF) regulates defense-related gene expression via GCC box and non-GCC box cis elements (Myb1 (GTTAGTT) and G-box(CACGTG));
Rth3	399	BOXCPSAS1	CTCCCAC	399	G→A	Box C in pea (P.s.) asparagine synthetase (AS1) gene; AS1 is negatively regulated by light; Box C binds with nuclear proteins, which was completed by a putative repressor element RE1.
Rth3	436	DOFCOREZM	AAAG	438	A→G	Core site required for binding of Dof proteins in maize (Z.m.);Dof proteins are DNA binding proteins, with presumably only one zinc finger, and are unique to plants; Four cDNAs encoding Dof proteins, Dof1, Dof2, Dof3 and PBF, have been isolated from maize; PBF is an endosperm specific Dof protein that binds to prolamin box; Maize Dof1 enhances transcription from the promoters of both cytosolic orthophosphate kinase (CyPPDK) and a
Rth3	600	BIHD1OS	TGTCA	600	T→C	Binding site of OsBIHD1, a rice BELL homeodomain transcription factor;
Rth3	600	WRKY71OS	TGAC	600	T→C	A core of TGAC-containing W-box" of, e.g., Amy32b promoter; Binding site of rice WRKY71, a transcriptional repressor of the gibberellin signaling pathway; Parsley WRKY proteins bind specifically to TGAC-containing W box elements within the Pathogenesis-Related Class10 (PR-10) genes.
Rum1	118	SURECOREATSULTR11	GAGAC	118	G→C	Core of sulfur-responsive element (SURE) found in the promoter of SULTR1;1 high-affinity sulfate transporter gene in Arabidopsis; SURE contains auxin response factor (ARF) binding sequence (GAGACA); its complementary seq. is GAGACA), and this core sequence is a part of it; this core seq. is involved in -S response;

Supplementary Table 2: Association analysis results

SNP	Marker	p-value	q-value	Trait	Allele	Allele Effect
RTCL_SNP5	290	0.0112	0.0275	sTRL	T/A	-9.36
RTCL_SNP6	296	0.0112	0.0275	sTRL	A/G	-9.36
RTCL_SNP7	298	0.0112	0.0275	sTRL	A/T	-9.36
RTCL_SNP9	317	0.0501	0.0587	sTRL	C/T	-5.63
RTCL_SNP11	324	0.0042	0.0197	sTRL	C/G/T	-7.24, -15.11
RTCL_SNP13	357	0.0254	0.0482	sTRL	A/T/G	9.72, 18.93
RTCL_SNP14	373	0.0090	0.0275	sTRL	G/A	-9.46
RTCL_SNP16	413	0.0005	0.0074	sTRL	A/C	11.75
RTCL_SNP17	468	0.0020	0.0151	sTRL	T/A/C	2.32, -9.42
RTCL_SNP18	473	0.0004	0.0074	sTRL	G/A	11.66
RTCL_SNP20	510	0.0203	0.0423	sTRL	T/A	7.45
RTCL_SNP21	531	0.0316	0.0515	sTRL	G/C	-8.75
RTCL_SNP22	543	0.0098	0.0275	sTRL	A/G	-9.26
RTCL_SNP23	547	0.0004	0.0074	sTRL	G/T	12.06
RTCL_SNP24	554	0.0065	0.0270	sTRL	T/C	9.09
RTCL_SNP26	597	0.0098	0.0275	sTRL	G/T	-9.26
RTCL_SNP30	632	0.0098	0.0275	sTRL	G/A	-9.26
RTCL_SNP34	703	0.0027	0.0151	sTRL	C/G	-11.35
RTCL_SNP39	720	0.0247	0.0482	sTRL	A/T	7.94
RTCL_SNP40	736	0.0027	0.0151	sTRL	C/A	14.32
RTCL_SNP44	799	0.0306	0.0515	sTRL	T/G/A	-9.71, -7.8
RTH3_SNP47	993	0.0143	0.0332	sTRL	G/A	-6.90
RTH3_SNP48	1010	0.0370	0.0515	sTRL	G/T	7.22
RTH3_SNP49	1064	0.0370	0.0515	sTRL	C/T	7.22
RTH3_SNP50	1181	0.0485	0.0587	sTRL	T/G	-8.28
RTH3_SNP51	1223	0.0485	0.0587	sTRL	A/G	-8.28
RTH3_SNP52	1229	0.0485	0.0587	sTRL	A/G	-8.28
RTH3_SNP53	1247	0.0029	0.0151	sTRL	C/G	-8.46
RTH3_SNP54	1268	0.0484	0.0587	sTRL	A/G	6.33
RTH3_SNP55	1295	0.0370	0.0515	sTRL	C/A	7.22
RTH3_SNP57	1322	0.0370	0.0515	sTRL	G/T	7.22
RTH3_SNP58	1349	0.0370	0.0515	sTRL	G/A	7.22
RTH3_SNP59	1430	0.0370	0.0515	sTRL	T/C	7.22
RTH3_SNP60	1451	0.0021	0.0151	sTRL	A/G	-9.13
RUM1_SNP68	1845	0.0202	0.0423	sTRL	T/G/A	15.68, 23.19
RTCL_SNP5	290	0.0465	0.0660	sRDW	T/A	-2.50
RTCL_SNP6	296	0.0465	0.0660	sRDW	A/G	-2.50
RTCL_SNP7	298	0.0465	0.0660	sRDW	A/T	-2.50
RTCL_SNP9	317	0.0101	0.0444	sRDW	C/T	-2.76
RTCL_SNP10	320	0.0348	0.0660	sRDW	C/T	4.05
RTCL_SNP11	324	0.0131	0.0444	sRDW	C/G/T	-1.87, -4.38
RTCL_SNP15	374	0.0202	0.0560	sRDW	G/A	-2.84
RTCL_SNP17	468	0.0329	0.0660	sRDW	T/A/C	-4.29, -6.04
RTCL_SNP18	473	0.0245	0.0616	sRDW	G/A	2.32
RTCL_SNP22	543	0.0144	0.0444	sRDW	A/G	-3.12
RTCL_SNP23	547	0.0383	0.0660	sRDW	G/T	2.35
RTCL_SNP26	597	0.0144	0.0444	sRDW	G/T	-3.12
RTCL_SNP30	632	0.0144	0.0444	sRDW	G/A	-3.12
RTCL_SNP34	703	0.0120	0.0444	sRDW	C/G	-3.35
RTCL_SNP40	736	0.0052	0.0444	sRDW	C/A	4.46
RTCL_SNP44	799	0.0062	0.0444	sRDW	T/G/A	-3.68, -0.91
RTH3_SNP60	1451	0.0077	0.0444	sRDW	A/G	-2.71

RUM1_SNP63	1661	0.0503	0.0660	sRDW	G/C	-2.41
RUM1_SNP68	1845	0.0314	0.0660	sRDW	T/G/A	6.23, 8.1
RTCL_SNP16	413	0.0371	0.1944	tTRL	A/C	16.97
RTCL_SNP18	473	0.0454	0.1944	tTRL	G/A	15.81
RTCL_SNP21	531	0.0345	0.1944	tTRL	G/C	-21.68
RTCL_SNP23	547	0.0325	0.1944	tTRL	G/T	17.43
RTCL_SNP24	554	0.0330	0.1944	tTRL	T/C	17.41
RUM1_SNP61	1606	0.0459	0.1944	tTRL	T/C	-16.02
RUM1_SNP65	1794	0.0093	0.1944	tTRL	T/A	22.81
RTCL_SNP10	320	0.0486	0.3014	tRDW	C/T	6.44
RTCL_SNP15	374	0.0391	0.3014	tRDW	G/A	-4.79
RUM1_SNP61	1606	0.0349	0.3014	tRDW	T/C	-4.88
RUM1_SNP65	1794	0.0098	0.3014	tRDW	T/A	5.73
RUL1_SNP73	1997	0.0282	0.3014	tRDW	A/G	7.03
RUL1_SNP74	2050	0.0333	0.3014	tRDW	C/T	6.75
RUL1_SNP75	2076	0.0333	0.3014	tRDW	G/A	6.75
RTCL_SNP20	510	0.0328	0.0592	fTRL	T/A	26.91
RTCL_SNP24	554	0.0246	0.0488	fTRL	T/C	29.25
RTH3_SNP48	1010	0.0191	0.0421	fTRL	G/T	32.14
RTH3_SNP49	1064	0.0191	0.0421	fTRL	C/T	32.14
RTH3_SNP54	1268	0.0136	0.0421	fTRL	A/G	31.42
RTH3_SNP55	1295	0.0191	0.0421	fTRL	C/A	32.14
RTH3_SNP57	1322	0.0191	0.0421	fTRL	G/T	32.14
RTH3_SNP58	1349	0.0191	0.0421	fTRL	G/A	32.14
RTH3_SNP59	1430	0.0191	0.0421	fTRL	T/C	32.14
RUM1_SNP61	1606	0.0363	0.0601	fTRL	T/C	-25.33
RUM1_SNP65	1794	0.0069	0.0421	fTRL	T/A	35.16
RUM1_SNP68	1845	0.0094	0.0421	fTRL	T/G/A	122.43, 140.12
RTH3_SNP47	993	0.0441	0.1609	fRDW	G/A	-5.34
RTH3_SNP48	1010	0.0102	0.0740	fRDW	G/T	8.71
RTH3_SNP49	1064	0.0102	0.0740	fRDW	C/T	8.71
RTH3_SNP54	1268	0.0245	0.1137	fRDW	A/G	7.04
RTH3_SNP55	1295	0.0102	0.0740	fRDW	C/A	8.71
RTH3_SNP57	1322	0.0102	0.0740	fRDW	G/T	8.71
RTH3_SNP58	1349	0.0102	0.0740	fRDW	G/A	8.71
RTH3_SNP59	1430	0.0102	0.0740	fRDW	T/C	8.71
RUM1_SNP61	1606	0.0130	0.0740	fRDW	T/C	-7.55
RUM1_SNP63	1661	0.0183	0.0933	fRDW	G/C	-8.24
RUM1_SNP65	1794	0.0130	0.0740	fRDW	T/A	8.26
RUM1_SNP68	1845	0.0088	0.0740	fRDW	T/G/A	30.35, 34.5
RUM1_SNP69	1901	0.0333	0.1416	fRDW	G/C	-7.46
RUM1_SNP70	1924	0.0503	0.1714	fRDW	T/C	-7.82
RUL1_SNP78	2379	0.0404	0.1587	fRDW	T/A/G	8.30, 10.97