bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted October 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Title

Epistatic genetic architecture of root length in *Arabidopsis thaliana*

Authors

Jennifer Lachowiec[1,2], Xia Shen[3,4,5], Christine Queitsch[1*], Örjan Carlborg[3*]

Affiliations

[1]Department of Genome Sciences, University of Washington, USA

[2]Molecular and Cellular Biology Program, University of Washington, USA

[3]Department of Clinical Sciences, Swedish University of Agricultural Sciences, Uppsala, Sweden

[4]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

[5]MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK

* Co-corresponding authors

Abstract

Efforts to identify loci underlying complex traits generally assume that most genetic variance is additive. This is despite the fact that non-additive genetic effects, such as epistatic interactions and developmental noise, are also likely to make important contributions to the phenotypic variability. Analyses beyond additivity require additional care in the design and collection of data, and introduce significant analytical and computational challenges in the statistical analyses. Here, we have conducted a study that, by focusing on a model complex trait that allows precise phenotyping across many replicates and by applying advanced analytical tools capable of

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted October 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

capturing epistatic interactions, overcome these challenges. Specifically, we examined the genetic determinants of *Arabidopsis thaliana* root length, considering both trait mean and variance as phenotypes. Estimation of the narrow- and broad-sense heritabilities of mean root length found that only the non-additive variance was significantly different from zero. Also, no loci were found to contribute to mean root length using a standard additive model based genome-wide association analysis (GWAS). We could, however, identify one locus regulating developmental noise (root length variance) and seven loci contributing to root-length mean through epistatic interactions. Four of the epistatic loci were also experimentally confirmed. The locus associated with root length variance contains a candidate gene that, when mutated, appears to decrease developmental noise. This is particularly interesting as most other known noise regulators in multicellular organisms increase noise when mutated. A mutant analysis of candidate genes within the seven epistatic loci identified four genes that affected root development, including three without previously described root phenotypes. In summary, we identify several novel genes affecting root development, demonstrate the benefits of advanced analytical tools to study the genetic determinants of complex traits, and show that epistatic interactions can be a major determinant of complex traits in natural *A. thaliana* populations.

Author summary

Complex traits, such as many human diseases or climate adaptation and production traits in crops, arise through the action and interaction of many genes and environmental factors. Classic approaches to identify contributing genes generally assume that these factors act additively. Recent methods such as genome-wide association studies often adhere to this additive genetics paradigm. However, additive models of complex traits do not reflect that genes can interact non-additively. In this study, we use *Arabidopsis thaliana* to determine the additive and non-additive

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted October 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

contributions to the complex trait root length. Surprisingly, much of the observed phenotypic variation in root length across genetically divergent strains was explained by genetic interactions. We mapped eight such non-additive loci and validated four epistatic genes using mutant analysis. For three of these, this is their first implication in root development. Together, our results emphasize the importance of considering both non-additive and additive effects when dissecting complex traits.

Introduction

Identifying the loci underlying quantitative phenotypes is among the central challenges in genetics. The current quantitative genetics paradigm is based on the assumption of additive gene action, despite an increasing body of evidence showing that non-additive effects are crucial to many, if not most, biological systems [1–3]. The key arguments for remaining within the additive paradigm are that many genetic architectures with non-additive gene action display considerable additive genetic variance in populations [4] and that additive model-based approaches have facilitated detection of thousands of loci associated with many complex traits [5]. The focus on additive variation alone, however, can leave a considerable amount of genetic variance unexplored [1,6]. Therefore, to fully dissect a complex trait it is often necessary to explore alternative approaches that capture non-additive variation. Two biological mechanisms underlying non-additive variation in populations are developmental noise and epistatic interactions. Mounting evidence indicates that developmental noise, defined as stochastic phenotypic variation in the absence of genetic and environmental variation [7], can be genetically determined [2,8–12]. Statistical epistatic interactions between loci are pervasive in human and model organism complex traits whenever attempts are made to identify them [3,6].

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted October 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

To obtain a more complete compilation of the loci contributing to the variation of a complex trait, one therefore needs to consider developmental noise and epistatic interactions in addition to additive effects. A weakness of most statistical epistatic analyses completed in higher organisms is that few have been subjected to independent follow-up studies to replicate the epistatic associations, few have been able to illustrate the value of advanced analyses for identifying novel genes contributing to the studied phenotypes, and rarely have any have identified the functional mechanisms leading to the original statistical epistatic signals. Such work,however, is of utmost importance, both for identifying the most promising analytical approaches for future work and to identify candidate, epistatic genes for in-depth explorations of the biological mechanisms leading to non-additive genotype-phenotype relationships.

Genome wide association studies (GWAS) have emerged as an effective method for mapping loci underlying phenotypic variation for complex traits. However, GWAS is still under-utilized in efforts to understand the genetic underpinnings of developmental noise and epistatic interactions. Thus far, genes that regulate developmental noise in morphology and gene expression levels or epistatic interactions have been primarily identified with mutant screens, candidate gene approaches [2,3,7,12–19], and QTL mapping [8–10,20–22]. Therefore, a large portion of the genome is typically neglected, in which additional regulators of developmental noise and epistatic interactions are likely to exist.

There are two major challenges in using GWAS to identify non-additive variation: the lack of sufficiently comprehensive experimental datasets and the computational complexity of unbiased analyses for multi-locus epistasis. To map regulators of developmental noise using GWAS, the trait in question needs to be amenable to high-throughput quantitation and must exhibit a measurable amount of developmental noise. The root length of *Arabidopsis thaliana* seedlings is

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted October 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

an ideal trait to analyze for developmental noise because roots are easy to measure precisely across thousands of seedlings [10]. To map epistatic interactions, the GWAS method used must have sufficient power. It is generally thought that GWAS is underpowered to detect epistasis. This lack of power arises from the generally small number of observations in the multi-locus, minor-allele genotype classes and the need to use stringent, multiple-testing corrected significance-thresholds to account for the large number of statistical tests performed [23]. Recent efforts have shown that it is possible to overcome these challenges through extensive analyses utilizing large populations [24]. However, even in moderately sized populations interactions with large effects can be mapped, in particular, in inbreeding organisms with smaller genomes. In these species. the number of multi-locus genotype-classes is smaller and the phenotypes can be repeatedly measured for each genotype. This facilitates powerful epistatic GWAS analyses due to the availability of precise genotype-values and a lower multiple-testing penalty. *A. thaliana* has previously been used for standard GWAS for many traits [25,26], including several root phenotypes [27–29]. *A. thaliana* has a small genome (~125 Mb), large numbers of readily available inbred accessions, a large knowledge base on how to efficiently score quantitative phenotypes in large numbers of individuals and powerful experimental resources, such as collections of knock-out lines. It is therefore a particularly well-suited model for performing GWAS analyses of developmental noise and exhaustive scans for epistasis and evaluate whether non-additive genetic analyses are able to uncover novel contributions of genes to the studied phenotypes.

In this study, we completed GWAS scans for additivity, developmental noise, and for epistatic interactions for *A. thaliana* seedling root length. In the GWAS for additive loci, none were found to be associated with mean root length. This was consistent with the low observed narrow-sense

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted October 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

heritability. One locus was found to affect developmental noise in seedling roots. This locus contained a promising candidate gene potentially involved in modifying phytohormones or cell wall components. Only the broad-sense heritability of *A. thaliana* mean root length was significantly different from zero in the studied population, implying that the trait is primarily under non-additive genetic control. Our exhaustive scan for two-locus interactions affecting mean root length further supported this result as four genome-wide significant interacting pairs were found. Together, these epistatic pairs explained the majority of the non-additive variance underlying root length. Using mutant analysis, we explored candidate genes in the epistatic regions to identify the most likely contributors to the detected associations and reduce the number of genes to explore in future in-depth explorations of the mechanisms underlying the observed epistasis. Three of the six tested epistatic regions were shown to contain novel genes affecting root length, and a fourth region contained a gene that previously reported to affect root length [30]. Further analyses of available whole-genome re-sequencing data revealed candidate regulatory and coding variants within, or in LD with, these genes. Our results clearly illustrate the general importance of considering non-additive genetics to detect novel genes affecting complex traits and demonstrate a powerful computational genetics approach for use in future studies. They also provide several promising candidates for future functional studies to dissect the molecular mechanisms underlying genetic interactions and developmental noise shaping roots in natural *A. thaliana* populations.

Results

Epistasis is the primary genetic determinant of phenotypic variation in seedling root length

To study the genetic regulation of mean root length and developmental noise (variance, σ^2), we generated a comprehensive data set containing the *A. thaliana* root mean lengths and variances

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted October 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

across 93 natural accessions (Figure S1, Table S1). The 93 accessions used here were previously genotyped at ~215,000 loci and used for standard GWAS for many plant phenotypes [25]. We generated this new data set because previous, publicly available data were created to estimate trait means and hence were not suitable for studying developmental noise [7,8,25]. Root developmental noise for each accession was independently calculated from three biological replicates with n = 70 (Figure S1). As *A. thaliana* is an inbreeding species, uncontrolled genetic variance is virtually eliminated by such extensive replication, providing very precise genotype values for use in a GWAS analysis. We grew seedlings in a randomized design on standard plant medium in temperature-controlled chambers to reduce environmental variation among individuals [10]. This experimental design allows for parsing of genetic and environmental factors from developmental noise underlying root length [7,31]. We examined the reproducibility of root developmental noise and found no significant differences among the three replicates for each accession (Figure 1a, ANOVA, $p = 0.079$). This reproducibility, together with the fact that root developmental noise across genotypes (accessions) varied significantly ($p = 7.5 \times 10^{-239}$, Brown-Forsythe test), suggests that root developmental noise has a genetic basis that could be dissected using association mapping.

We also estimated the pooled variances by combining the measurements from the three biological replicates for each accession after correcting for replicate effects (see Methods) to estimate accession-specific root developmental noise with even greater accuracy. We found a strong correlation between the mean noise of the three biological replicates and the pooled noise for each accession ($p = 1.8 \times 10^{-68}$, $R^2 = 0.97$, Figure S2a) and proceeded to use the pooled noise measurements in all further analyses.

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted October 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Several previous studies of loci associated with developmental noise found that many of these also affect trait means [8,10,32]. This potentially confounding relationship between trait mean and variance is a concern in examining developmental noise [31] as classical statistical models operate on the assumption that mean and variance are independent. If this assumption is violated, other models must be used. In our dataset, mean root length also varied significantly across accessions ($p = 0$, ANOVA, Figure 1b). We confirmed that mean and developmental noise of root length were not correlated ($p = 0.11$, $R^2 = 0.02$, Figure S2b). The genetic architectures underlying mean root length and noise can therefore be explored as independent traits in this study, and consequently any potential overlaps in the revealed architectures are independently derived.

To further determine whether the observed variation in mean root length and noise (measured as within-accession variance) across these accessions was under genetic control, and not purely stochastic, we estimated the heritability of both traits. To do so, we used a linear mixed model that simultaneously considers the mean and the dispersion of the data while correcting for population structure among the accessions implemented in the R/hglm package [33]. The broad sense heritability was 0.24 (95% CI 0.10 – 0.38) for mean root length and 0.0016 (95% CI 0 – 0.0026) for root length noise (Table 1). We further partitioned the genetic variance explained into additive and epistatic variance. For mean root length, almost all the variance was explained through epistatic effects (0.22) and only a small, insignificant fraction by additive effects (0.01; $p = 0.79$). For root length noise, both the additive (0.0011; $p = 0.10$) and epistatic (0.00050; $p = 0.47$) contributions were very little. These heritability patterns suggested that GWAS analysis would be unlikely to uncover loci explaining major portions of phenotypic variation in root

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted October 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

length mean and root developmental noise and that analyses only considering additive genetic effects on the mean root length would be inadequate.

No additive GWAS associations were detected for mean root length

Two GWAS methods were used to screen for additive loci affecting mean root length. The first method was a standard additive model-based approach, with mixed-model-based correction for population structure, as implemented in the R-package GenABEL [34]. We accounted for population structure by using the principle components of the genomic kinship matrix ("egscore" procedure in R/GenABEL), based on the ~215,000 genotyped SNPs for the 93 accessions across the entire *A. thaliana* genome. The second method was based on a whole-genome generalized ridge regression heteroscedastic effects model (HEM), in which all SNPs were included simultaneously and their contributions estimated as random effects using the R/bigRR package [35]. A 5% genome-wide significance threshold was determined by permutation testing. None of the tested SNPs were found to be significantly associated with mean root length in either of these analyses (Figure S3a, b). Given the low estimate for the narrow-sense heritability for mean root length in this population ($h^2 = 0.01$), this outcome was unsurprising.

Root length variance is associated with a SNP in the gene At3g46700

Next, we performed GWAS analyses to map loci affecting root developmental noise (*i.e.* within-accession root length variance, σ^2), using the two same GWAS models used in the analyses of root length mean. No SNP was significantly associated with the trait using the standard additive GWAS model (Figure S4), however, one SNP was just below the conservative Bonferroni-corrected significance threshold ($p = 1.61 \times 10^{-7}$, threshold = 2.34×10^{-7}). Using the HEM GWAS model, the minor allele for this SNP (MAF = 0.17) on chromosome 3 at position 17,201,307bp (rs346930686 abbreviated here as 3_17201307) was significantly associated with

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted October 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

greater developmental noise in root length (|SNP effect| = 2.26×10^{-3}, threshold = 2.04×10^{-3}, Figure 2a, b). Surprisingly, this mapped SNP explains 28.9% of the variance of the developmental noise, far exceeding the estimate of the narrow-sense heritability estimated under the assumption of a polygenic trait architecture. This finding suggests that root developmental noise has a relatively simple genetic architecture (see Discussion).

The SNP 3_17201307 encodes a missense variant (A -> E) in the gene At3g46700 predicted to be tolerated by SIFT [36]. We explored whether this associated SNP was in linkage disequilibrium (LD) with any other SNPs ± 15 kb away—LD typically decays within 10kb [37] in *A. thaliana*. The associated SNP was in LD ($R^2 > 0.8$) with ten other SNPs identified in the 1001 Genomes Project [38] within this 30 kb window. Nine of these SNP affected the coding or upstream non-coding region of At3g46700. At3g46700 is a member of the UDP-glycosyltransferase superfamily based on sequence homology [39]. UDP-glycosyltransferases catalyze the enzymatic addition of sugars onto a variety of other molecules, such as secreted components of the plant cell wall [40] and phytohormones [41–45]. Thus far, there are no known substrates of the enzyme encoded by At3g46700. At3g46700 belongs to a cluster of tandemly replicated UDP-glycosyltransferases, of which only At3g46700 and one other, At3g46690, show chromatin accessibility in their regulatory regions in root tissue [46]. One SNP in LD with the leading SNP lies in the coding region of At3g46690, causing a S->T missense change, predicted to be tolerated by SIFT [36]. Both At3g46700 and At4g46690 are most highly expressed in seedling and adult roots compared to other tissues, supporting these genes as strong candidates for the observed association [47].

Further explorations of the natural variation present in At3g46700 and At3g46690 among the 93 accessions using data from the 1001 Genomes Project [38] revealed the presence of two stop

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted October 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

codons in At3g46700 (Chromosome 3 at 17,200,906 and 17,201,617 bp respectively; SNPs 3_17200906 and 3_17201617) in 4 and 21 accessions, respectively. Notably, these two stop codons are not in LD with the leading, trait-associated SNP. To explore whether these stop codons were associated with root developmental noise, we compared accessions containing a stop codon to the accessions without stop codons. Root length noise was indeed greater among accessions with a functional At3g46700, but not significantly so (Figure 2c, $p = 0.06$, t-test, $n = 63$). As few accessions carried these mutations, further analyses are needed to fully explore whether functional At3g46700 decreases root length noise.

An epistatic GWAS analysis uncovered multiple interacting loci contributing to mean root length

To identify loci contributing to the large epistatic genetic variance for mean root length, we performed an exhaustive two-locus SNP-by-SNP epistasis analysis (PLINKv1.07) [48]. Despite the limited population size, several factors increased our chances of finding significant interactions: the large contribution of non-additive genetics (Table 1), the high precision in estimating the phenotypic values due to the extensive replication, the presence of only four two-locus genotypes due to inbreeding, and a reduced number of tested pairwise combinations due to small genome size. To further enhance power, we did not test for epistasis for pairs of SNPs where the minor two-locus genotype class contained few observations. We achieved this objective by applying two data filters: first, only SNPs with a minor allele frequency greater than 25% were included, and second all SNP-by-SNP combinations with fewer than four accessions in the minor genotype class were removed. In the epistatic analyses, population stratification was accounted for by performing the GWAS on the residuals from a linear mixed model including the genomic kinship matrix, as described above for the single-locus GWAS analyses. For

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted October 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

significance testing, we used a Bonferroni-corrected threshold calculated by estimating the number of independent pair-wise tests for the number of linkage blocks across the *A. thaliana* genome [37]. In total, the epistatic analysis included approximately 78 million independent tests. Six SNP-pairs, representing seven genomic locations and four unique combinations of loci, were significantly associated with mean root length after accounting for multiple testing (Figure 3a, Table 2, Figure S5). Two pairs were detected twice by associations to tightly linked SNPs located 1.5 kb and 505 bp apart (Chromosome 3 at 9,272,294 and 9,273,674 bp - SNPs 3_9272294/ 3_9273674 - and Chromosome at 15,862,026 and 15,862,525 bp - SNPs 5_15862026/5_15862525, respectively ; Table 3), and one SNP (Chromosome 3 at 66,596 bp - SNP 3_66596) was part of two unique interacting pairs. The other two SNP-pairs contained independent SNPs ($R^2 < 0.8$).

We found that the four unique SNP pairs together explained 59% of the epistatic variance in mean root length (Table 1). We further examined mean root lengths and genotype frequencies in each two-locus genotype class for the four significant SNP pairs (Figure 3b) and observed the same pattern for all pairs: the combination of the two minor alleles was associated with the longest mean root length. Although this type of epistasis is expected to also result in a marginal additive genetic effect, the low frequency of the two-locus genotypes associated with long roots most likely explains why their contribution to the additive genetic variance is too low to allow detection by standard GWAS analysis.

To explore these significant epistatic associations further, we extracted 13 million SNPs obtained by whole-genome re-sequencing in the 1001 Genomes Project [49]. SNPs that were in LD ($R^2 >$ 0.8) and located within 5kb upstream or downstream of the leading SNP were considered further

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted October 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(Table 2). For each pair, the predicted functional effects of the SNPs that fulfill these criteria, together with their locations, are listed in Table 3. In total, 13 genes and two transposons were found in LD with the leading SNPs in the seven regions associated with mean root length through epistatic interactions. We present the results from our experimental evaluation of selected candidate genes below (Table 2); additional results are presented in the Supplementary Text. In short, we assessed whether likely loss-of-function mutants of these genes significantly alter root length. For three of these (*NAC6*, *At3g28880* and *TFL1*), we are the first to show that they contribute to root growth, and for one (*ATL5*) our results confirm previous findings.

Genetic analysis of candidate genes identifies four genes involved in root development

ATL5/OLI5/PGY3 insertion mutants decreased root length

Two linked SNPs on chromosome 3 (3_9272294 and 3_9273674; Table 2) were associated with *A. thaliana* root length through an epistatic interaction with a distal SNP on the same chromosome (3_66596). Twenty-one SNPs located in four genes were in high LD with the two linked SNPs (Table 3; Table 4). For three of these (At3g25520, At3g25530 and At3g25540), we were able to obtain insertion mutants and scored their mean root length compared to wild-type controls. The insertion mutant for At3g25520 had significantly shorter roots than wild type ($p = 1.2 \times 10^{-4}$; Tukey's post-hoc test); the insertion mutants for At3g25530 and At3g25540 did not differ significantly ($p = 1.00$ and $p = 0.23$, respectively) from wild type (Figure 4).

At3g25520 encodes ATL5/OLI5/PGY3, a 5S rRNA binding protein whose promoter is highly accessible in seedling roots [46], corresponding to expression in roots [47]. Our finding agrees with earlier studies, in which the loss-of-function mutant *oli5-1* exhibits shorter primary roots

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted October 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

than wild-type [30]. In the tested accessions, three SNPs were found in the *ATL5* promoter and a fourth in an intron, suggesting variable transcriptional regulation across accessions.

No insertion mutant was available for the gene At3g25545. This gene has not been reported to affect root length, but it is known to be expressed in the vasculature of seedling roots and imbibed seeds [47]. In the tested accessions, this gene contains three SNPs in its putative promoter, motivating future studies to explore its potential role in root growth.

In summary, our results support that the GWAS association between the SNPs 3_9272294/3_9273674 and root length is due to polymorphisms affecting the function or transcription of the gene *ATL5*. The existence of several polymorphisms in the *ATL5* promoter region implies altered transcriptional control as the most likely explanation.

A novel role for NAC6 in the regulation of root length

Two linked SNPs on chromosome 5 (5_15862026 and 5_15862525; Table 2) interacted with an SNP on chromosome 1 (17,257,526 bp - SNP 1_17257526; Table 2). Both of these linked SNPs, as well as the other SNPs in high LD, were intergenic with the nearest gene, *NAC6* (At5g39610; Table 4), which is located 1.9 kb away. *NAC6* is a transcription factor regulating leaf senescence [50] and is highly expressed in senescing leaves and in maturing seeds [47]. We phenotyped the available At5g39610 insertion mutant and found a significant decrease in mean root length ($p = 5.82 \times 10^{-10}$; Figure 4), strongly implicating *NAC6* in the regulation of root length. Further studies are needed to explore how this gene interacts epistatically with its paired locus, which resides within a transposon on chromosome 1 (see below). Both loci coincide with accessible chromatin, which is a hallmark of regulatory DNA, across several tested accessions [46]. We

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted October 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

speculate that the polymorphisms in our tested accessions alter the regulatory potential of these regions.

The previously uncharacterized gene At3g28880 contributes to root length

A third locus on chromosome 3 (10,891,195 - SNP 3_10891195; Table 2) interacted with a locus on chromosome 5 (1,027,939 bp - SNP 5_1027939; Table 2). SNPs in LD with 3_10891195 were found in three genes: At3g28865, At3g28870, and At3g28880 (Table 4). None of these genes previously had been associated with root length. By examining available insertion mutants, we established a novel role for the gene At3g28880 in root development; its mutant showed significantly decreased mean root length compared to wild-type controls ($p = 4.5 \times 10^{-10}$; Figure 4). At3g28880 encodes a previously uncharacterized ankyrin family protein. There is no information on developmental expression patterns for this gene; although the gene's putative regulatory region is accessible in seeds and seedling tissue [46]. Across the tested accessions, a number of promoter-proximal SNPs were found in At3g28880, possibly altering gene regulation.

A third candidate gene in the region, At3g28870, was previously uncharacterized. This gene contains a number of conserved domains, including a Sec63 domain that harbored a missense variant across the tested accessions [39]. Unfortunately, we were unable to obtain an insertion mutant for this gene, but in light of the validated effect of At3g28880 on mean root length, we consider At3g28870 a less likely candidate for the observed GWAS association.

Mutant TFL1 increases mean root length

The SNP 5_1027939 was detected via its epistatic interaction with the At3g28880 locus on chromosome 3 (Table 2). This polymorphism and the only other polymorphism in LD with it, are both intergenic between the genes At5g03840 and At5g03850 (Table 4). Phenotyping the

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted October 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

insertion mutants for these genes revealed that the mutant for At5g03840, which encodes *TFL1*, showed significantly longer roots than wild-type controls ($p = 5.7 \times 10^{-10}$; Figure 4), whereas no significant phenotypic effect was found for the At5g03850 mutant. Interestingly, these results implicate *TFL1* as a repressor of root growth.

Earlier work supports At5g03840, *TERMINAL FLOWER 1* (*TFL1*), as a highly probable gene to be involved in epistatic interactions as it affects several plant traits. For example, *TFL1*'s role in floral initiation and morphology are well-established in many plant species [51]. It also plays a role in regulating protein storage in vegetative tissues, such as roots or seeds [52]. *TFL1* is highly expressed in both adult and seedling roots [47]. Our finding that *TFL1* controls root length, and possibly does so via interactions with At3g28880 in natural populations, is particularly poignant as a previous study reported no root length defects for a different *tfl1* mutant with multiple other pleiotropic phenotypes [53]. Our finding extends the functional reach of this multi-functional regulator and implies that its role in root development needs to be studied in the context of epistatic interactions.

Although we detected no phenotypic effect on root length for a mutant of At5g03850, which encodes a nucleic acid-binding, OB-fold-like protein, we cannot rule out that it may play some role in root development as it is expressed in seedling roots [47]. Our findings, however, strongly implicate *TFL1* as the most likely gene causing the GWAS association in the tested accessions.

Discussion

In this study, we set out to identify genes underlying seedling root length mean and variance (as a measurement of developmental noise). We found that the genetic architecture of mean root

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted October 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

length differed greatly from that of root length variance in the studied collection of natural *A. thaliana* accessions. The narrow-sense heritability was not different from zero for either trait, whereas there was significant broad-sense heritability for mean root-length (Table 1). The genetics underlying developmental noise (variance; Figure 1) can therefore apparently not be captured using standard heritability estimation, whereas a significant epistatic contribution to mean root length is observed. Notably, we observed a very large contribution of non-additive effects to the phenotypic variance of mean root length. With these findings in hand, we completed a comprehensive set of GWAS scans to identify additive and epistatic loci affecting mean root length as well as loci affecting developmental noise expressed as root length variance.

Consistent with the low narrow-sense heritability for mean root length, we were unable to identify any additive loci that reached the Bonferroni-corrected significance threshold in our analyses. In contrast, we identified a locus that was significantly associated with root developmental noise in spite of this trait's non-significant narrow-sense heritability. The associated marker explains 28.9% of the variance of the developmental noise. Although this high value might seem surprising, it actually indicates that root developmental noise has a relatively simple genetic architecture. The heritability estimation assumes a polygenic model, in which the genomic kinship matrix is fitted to capture the total heritability. In this type of estimation, it is assumed that the average genomic relationship between individuals is a measure of the shared genetic architecture of the trait. This averaging is appropriate when a large number of loci affect the trait and where each of these explains a small fraction of the genetic variance. However, if the genetic architecture is due to one or a few genes with large effects, such as for root developmental noise, the kinship matrix estimated based on all markers across the genome will be a poor representative for the true architecture and capture little of the genetic variance for the

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted October 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

trait. Therefore, our results indicate that the mapped locus is a major locus for root developmental noise, and that the other parts of the genome explain little in addition to it.

By taking potential epistatic interactions shaping mean root length into account, we identified seven loci involved in four unique epistatically interacting pairs of loci that together explained a majority of the total estimated epistatic genetic variance (Table 2). Consequently, in this study we were able to attribute much of the epistatic genetic variance in a complex trait to the effects of individual loci. This result stands in stark contrast to most other studies of complex traits and illustrates the benefits of taking an unbiased modeling approach to dissect a complex trait.

GWAS have earlier been used to identify loci affecting developmental noise in gene expression [1,8,21]. In this study, we performed the first GWAS for developmental noise of a morphological trait. We assessed a large number of individuals, in replicate experiments, to estimate trait variance accurately. One locus was found to affect root length variance. The associated SNP, and SNPs in LD, affect At3g46700 and At3g46690, genes that encodes UDP-glycosyltransferases. These genes are members of a family of enzymes that add sugar groups to a variety of substrates [54]. Several potential substrates of At3g46700 and At3g46690 are known to be important for root growth, including cell wall components and phytohormones [40–45]. The 25 kb region flanking At3g46700 and At3g46690 also encodes five additional duplicated copies of UDP-glycosyltransferases with the same exon-intron structure as At3g46700 and At3g46690. At3g46700 shares 59-73% amino acid identity with these other UDP-glycosyltransferases [55]. The presence of closely related, and potentially partially redundant, copies of this UDP-glycosyltransferase may explain the tolerance of the stop codon present in At3g46700 in many

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted October 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

accessions with regard to general effects on root development. In fact, accessions carrying the early stop codons in At3g46700 showed wild-type mean root length ($p = 0.41$, t-test, unequal variances), suggesting that At3g46700 specifically affects root variance without affecting trait mean.

We suggest that At3g46700 translates stochastic fluctuations in external or internal micro-environments in root development, resulting in root length variation among isogenic siblings. If so, loss of At3g46700 function would result in less variance in root length. Indeed, we observed that the root length variance tended to decrease in accessions with the early stop codons in At3g46700. This scenario is reminiscent of an earlier study by Hall et al. (2007), which reported that an *erecta*-mutant allele decreased developmental noise in the stem length of *A. thaliana* seedlings. It is generally assumed that functional redundancy among gene family members constrains developmental noise and stabilizes phenotypes [56]. Our findings challenge the generality of this assumption, as stop codons in At3g46700 were associated with decreased developmental noise, implying the existence of alternative mechanisms to stabilize traits. Our results are also consistent with the speculation that At3g46700 function in promoting developmental noise may be dispensable under certain circumstances, such as in comparatively stable environments.

Functional dissection of epistasis is a daunting task in natural populations. A first step in taking on this challenge is to reduce the list of potential candidate genes and mutations for further, in-depth functional explorations. Here, we use mutational analysis of the genes that harbor SNPs in LD with the associated SNPs in the epistatic GWAS analysis to identify the most likely candidate genes for the inferred associations. We tested six of the seven epistatic regions with available insertion mutants. For four loci, we succeeded in identifying candidate genes affecting

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted October 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

root length, three of which were newly implicated in root development. This is an astounding success rate, particularly in the light of widespread redundancy underlying morphological traits in *A. thaliana*. The majority of single-gene mutations affects neither trait means nor trait variance [2,7,57]. Our results demonstrate the benefits of epistatic analyses to identify novel genes involved in complex traits and provide a basis for future detailed studies of the complex gene networks involved in root development.

The novel genes implicated in shaping root length here were functionally diverse. For example, the transcription factor *NAC6* has been previously described as a positive regulator of plant senescence, increasing in expression in aging rosette leaves [58]. We found that natural variation in the *NAC6* promoter was associated with root length through an interaction with a transposon-rich region on chromosome 1 and confirmed that *NAC6* affected root length in the Col-0 reference background. It is intriguing to speculate how regulatory variation in *NAC6* interacts with polymorphisms in transposons to affect root length. In fact, transposons have been previously found to influence expression phenotypes in *A. thaliana*. For example, a transposon present in the intron of *FLC* in some accessions alters *FLC* expression [59]. Further, a recent study found that the presence of polymorphisms in transposons is positively correlated with gene expression variation in adjacent genes [60]. The transposon-rich region is located on a different chromosome and not adjacent to *NAC6*. We speculate that the observed epistatic interaction arises indirectly with genes that are located adjacent to the transposon-rich region, but themselves do not contain trait-associated SNPs.

Similarly to *NAC6*, the trait-associated SNPs in the two other genes newly implicated in root development resided in putative regulatory regions, consistent with the established role of

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted October 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

expression variation in shaping phenotypic variation [25,61,62]. Whereas *TFL1* has been previously implicated in a variety of developmental phenotypes, our study provides the first functional description for *At3g28880*.

It will be a major endeavor to functionally dissect the observed epistatic associations. None of the genes affecting root development (*ATL5*, *NAC6*, *TFL1*, At3g28880) contain non-synonymous mutations with likely phenotypic effects. Also, none of these genes have previously been described to interact with their respective epistatic locus. Future functional studies will have to explore the effects of the observed regulatory SNPs in transgenic plants in several backgrounds to study the mechanisms underlying these interactions in detail. This endeavor will require significant time and resources, yet we believe that *A. thaliana* root length represents an excellent model trait for future in-depth studies of genetic interactions. Mean root length is here shown to be highly epistatic and regulated by strong interactions between a limited number of loci, making it feasible to dissect the genetic architecture underlying highly complex multi-locus epistatic GWAS associations.

Our study demonstrates the benefits of using a more analytically involved approach to dissect a complex trait in addition to using standard additive model-based GWAS analysis. The need for using an epistatic model was indicated by our analysis of both the narrow- and broad-sense heritabilities for mean root length. Here, our results clearly showed that we would miss most of the genetic variance if only using standard GWAS models. Indeed, the use of the epistatic model allowed us to infer several novel genes and reconfirm one other gene affecting root length. No significant associations were found using standard GWAS models. Similarly, for the analyses of root length variance, the more complex multi-marker HEM analysis was more powerful for

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted October 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

identifying associated loci when compared to standard GWAS. Future GWAS-based studies aiming to dissect the genetics of complex traits are therefore advised to consider more advanced multi-locus based approaches to improve the power to identify important new genes.

Materials and Methods

Phenotyping

In order to reduce environmental variation among accessions, eighteen individuals from each of the 93 accessions studied in Atwell et al (2010) (Table S1) were vernalized at 4° as seedlings for six weeks to synchronize growth and flowering. The five most developmentally advanced seedlings from each accession were then transferred to soil in a randomized design. Plants were grown in long-day conditions at 22°. Flats were rotated three times per week to reduce position effects on plant development. Seeds were collected over a period of three months as the plants dried. Equal numbers of seeds were pooled from 3-5 parent plants to reduce the environmental contribution of plant placement across flats of plants. Ethanol sterilized seeds were planted on 1× Murashige and Skoog (MS) basal salt medium supplemented with 1× MS vitamins, 0.05% MES (wt/vol), and 0.3% (wt/vol) phytagel in a semi-randomized design with n = 70 per accession. Four sets of twenty-three accessions plus a control accession (Col-0) made up a set. Each set was replicated three times providing the standard error for variance, with a total n = 210 planted for each accession. The seeds were stratified at 4° for three days and then grown for seven days in darkness with the plates in a vertical position. A photograph of each plate was taken, and root length was measured manually using the 'freehand' function in ImageJ1.46r [63]. Non-germination, missing organs, and delayed development were noted.

Data standardization

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted October 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Thirty-one seedlings with hypocotyls less than 5mm were removed because it is likely that germination was severely delayed [10]. Between-set differences were corrected by subtracting the difference between Col-0 in each replicate and set and the global mean for Col-0. Systematic differences between replicates were still present and corrected within each set by using the residuals from a model in which root length was explained by the within-set replicates.

Heritability estimation

Heritability of root length variance and mean were estimated using the repeated measures with genotype as a fixed effect using ANOVA. To parse the contribution of additive and epistatic effects, a linear mixed model including both additive and epistatic effects as random effects was fitted using the R/hglm package [33], i.e.

$$y = \mu + a + b + e,$$

where

$$a \sim N(0, G\sigma_a^2)$$

are the additive genetic effects, and G is the genomic kinship matrix;

$$b \sim N(0, G \circ G\sigma_b^2)$$

are the epistatic effects, and G is the genomic kinship matrix; *e* are independent and normally distributed residuals. The narrow sense heritability was estimated as the ratio of the additive genetic variance to the total phenotypic variance, and the broad sense heritability was estimated as the ratio of the sum of the additive and epistatic variance to the total phenotypic variance.

GWAS analysis

Using the transformed mean and variance root length estimates, and the genotypes for the 93 accessions from Atwell et al. (2010), we performed a series of analyses to detect marker-trait

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted October 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

associations. For the standard additive GWAS with population structure controlled, we used the function egscore from R/GenABEL [34]. The functions bigRR() and bigRR_update() in the R/bigRR package [35] were used to perform a GWAS analysis, in which all genome-wide SNP effects are modeled simultaneously as random effects, and the effects were estimated via the generalized ridge regression method HEM. The maximum absolute effect sizes in 1000 permutations were extracted to determine an empirical 5% genome-wide significance threshold (p = 0.0020).

To screen the genome for pairs of epistatic SNPs, we used the PLINK –epistasis procedure [48] that is based on the model:

$$Y \sim b_0 + b_{1.A} + b_{2.B} + b_{3.AB} + e,$$

which considers allele by allele epistasis in $b_{3.AB}$. In the analysis, all possible pairs of SNPs with a minor allele frequency > 25% were tested.

We filtered out the pairs where there were fewer than four accessions in the minor two-locus genotype-class. Further, all combinations in which the p-value was > 1×10^{-8} were also removed as it was considered unlikely that they would be significant after correction for population structure and applying a multiple-testing corrected significance threshold. For the remaining pairs, a linear mixed model including fixed, additive and epistatic effects, as in the PLINK based initial scan, were fitted together with a kinship correction for population stratification, as in the single locus analyses, using the function R/hglm. We derived a multiple-testing corrected significance threshold for this epistatic analysis ($p = 3.2 \times 10^{-10}$) by estimating the number of independent tests based on the number of estimated LD blocks in the genome. For this, we used the fact that the *A. thaliana* genome is approximately 125Mb with average LD-blocks of about 10kb (12,500; [37]), and then applied a Bonferroni correction for the 78 million

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted October 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

independent tests performed. Using data on the sixty-three accessions from the GWAS that were unambiguously identified in the 1001 Genomes data [38], we then identified additional SNPs in LD with the leading SNP using the function *LD* in the *genetics* package in R across a 10kb region around the marker.

Examination of SNP-associated genes

The Ensembl Variant Effect Predictor based on the TAIR10 release of the *A. thaliana* genome was used to determine the effects of the leading SNPs and the SNPs in high LD with them. Genes were considered as candidates if they were within 1kb of a variant. Expression of the candidate genes was determined using the BAR eFP Arabidopsis browser [47].

Validation of interacting loci

T-DNA lines were obtained for the candidate genes (Table S2). Root length was ascertained as described above (n = 20). Tukey's HSD post hoc test was used to compare root lengths between the T-DNA lines and the wild-type accession (Col-0).

Acknowledgments

We thank Mats Pettersson and Ronnie Nelson for useful discussions and Karla Schultz and members of the Queitsch lab for experimental assistance.

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted October 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

1. Nelson RM, Pettersson ME, Li X, Carlborg Ö (2013) Variance Heterogeneity in Saccharomyces cerevisiae Expression Data: Trans-Regulation and Epistasis. PLoS One 8: e79507.

2. Queitsch C, Carlson KD, Girirajan S (2012) Lessons from model organisms: phenotypic robustness and missing heritability in complex disease. PLoS Genet 8: e1003041. Available: http://www.ncbi.nlm.nih.gov/pubmed/23166511. Accessed 27 May 2014.

3. Mackay TFC (2014) Epistasis and quantitative traits: using model organisms to study gene-gene interactions. Nat Rev Genet 15: 22–33. Available: http://dx.doi.org/10.1038/nrg3627.

4. Hill WG, Goddard ME, Visscher PM (2008) Data and theory point to mainly additive genetic variance for complex traits. PLoS Genet 4: e1000008. Available: http://dx.plos.org/10.1371/journal.pgen.1000008. Accessed 6 August 2013.

5. Visscher PM, Brown MA, McCarthy MI, Yang J (2012) Five years of GWAS discovery. Am J Hum Genet 90: 7–24. Available: http://www.sciencedirect.com/science/article/pii/S0002929711005337. Accessed 11 July 2014.

6. Carlborg O, Haley CS (2004) Epistasis: too often neglected in complex trait studies? Nat Rev Genet 5: 618–625. Available: http://dx.doi.org/10.1038/nrg1407. Accessed 8 September 2013.

7. Lempe J, Lachowiec J, Sullivan AM, Queitsch C (2012) Molecular mechanisms of robustness in plants. Curr Opin Plant Biol. Available: http://www.ncbi.nlm.nih.gov/pubmed/23279801.

8. Jimenez-Gomez JM, Corwin JA, Joseph B, Maloof JN, Kliebenstein DJ (2011) Genomic analysis of QTLs and genes altering natural variation in stochastic noise. PLoS Genet 7: e1002295. Available: http://dx.plos.org/10.1371/journal.pgen.1002295. Accessed 21 August 2013.

9. Hall MC, Dworkin I, Ungerer MC, Purugganan M (2007) Genetics of microenvironmental canalization in Arabidopsis thaliana. Proc Natl Acad Sci U S A 104: 13717–13722. Available: http://www.ncbi.nlm.nih.gov/pubmed/17698961.

10. Sangster TA, Salathia N, Undurraga, Milo R, Schellenberg K, et al. (2008) HSP90 affects the expression of genetic variation and developmental stability in quantitative traits. Proc Natl Acad Sci U S A 105: 2963–2968. Available: http://www.ncbi.nlm.nih.gov/pubmed/18287065.

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted October 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

11. Doyle MR, Davis SJ, Bastow RM, McWatters HG, Kozma-Bognár L, et al. (2002) The ELF4 gene controls circadian rhythms and flowering time in Arabidopsis thaliana. Nature 419: 74–77.

12. Levy SF, Siegal ML (2008) Network hubs buffer environmental variation in Saccharomyces cerevisiae. PLoS Biol 6: e264. Available: http://www.ncbi.nlm.nih.gov/pubmed/18986213. Accessed 27 November 2013.

13. Rinott R, Jaimovich A, Friedman N (2011) Exploring transcription regulation through cell-to-cell variability. Proc Natl Acad Sci U S A 108: 6329–6334. Available: http://www.pnas.org/content/108/15/6329. Accessed 9 November 2013.

14. Mackay TF (2001) The genetic architecture of quantitative traits. Annu Rev Genet 35: 303–339. Available: http://www.annualreviews.org/doi/abs/10.1146/annurev.genet.35.102401.090633. Accessed 28 November 2013.

15. Rowe HC, Hansen BG, Halkier BA, Kliebenstein DJ (2008) Biochemical networks and epistasis shape the Arabidopsis thaliana metabolome. Plant Cell 20: 1199–1216. Available: http://www.plantcell.org/content/20/5/1199. Accessed 29 July 2014.

16. Pyun J-A, Kim S, Cha DH, Kwack K (2013) Epistasis between polymorphisms in TSHB and ADAMTS16 is associated with premature ovarian failure. Menopause (New York, NY).

17. Sapkota Y, Mackey JR, Lai R, Franco-Villalobos C, Lupichuk S, et al. (2013) Assessing SNP-SNP interactions among DNA repair, modification and metabolism related pathway genes in breast cancer susceptibility. PLoS One 8: e64896.

18. Huang W, Richards S, Carbone MA, Zhu D, Anholt RRH, et al. (2012) Epistasis dominates the genetic architecture of Drosophila quantitative traits. Proc Natl Acad Sci U S A 109: 15553–15559. Available: http://www.pnas.org/content/109/39/15553.full. Accessed 29 July 2014.

19. Vanhaeren H, Gonzalez N, Coppens F, De Milde L, Van Daele T, et al. (2014) Combining growth-promoting genes leads to positive epistasis in Arabidopsis thaliana. Elife 3: e02252. Available: http://elifesciences.org/content/3/e02252.abstract. Accessed 10 July 2014.

20. Carlborg Ö, Jacobsson L, Åhgren P, Siegel P, Andersson L (2006) Epistasis and the release of genetic variation during long-term selection. Nat Genet 38: 418–420.

21. Ansel J, Bottin H, Rodriguez-Beltran C, Damon C, Nagarajan M, et al. (2008) Cell-to-cell stochastic variation in gene expression is a complex genetic trait. PLoS Genet 4: e1000049.

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted October 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

22. Fu J, Keurentjes JJ, Bouwmeester H, America T, Verstappen FW, et al. (2009) System-wide molecular evidence for phenotypic buffering in Arabidopsis. Nat Genet 41: 166–167. Available: http://www.ncbi.nlm.nih.gov/pubmed/19169256.

23. Gibson G (2012) Rare and common variants: twenty arguments. Nat Rev Genet 13: 135–145. Available: http://www.nature.com/nrg/journal/v13/n2/full/nrg3118.html#B121. Accessed 24 May 2014.

24. Hemani G, Shakhbazov K, Westra H-J, Esko T, Henders AK, et al. (2014) Detection and replication of epistasis influencing transcription in humans. Nature.

25. Atwell S, Huang YS, Vilhjalmsson BJ, Willems G, Horton M, et al. (2010) Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature 465: 627–631. Available: http://www.ncbi.nlm.nih.gov/pubmed/20336072.

26. Filiault DL, Maloof JN (2012) A genome-wide association study identifies variants underlying the Arabidopsis thaliana shade avoidance response. PLoS Genet 8: e1002589. Available: http://www.plosgenetics.org/article/info:doi/10.1371/journal.pgen.1002589#pgen-1002589-g004. Accessed 28 July 2014.

27. Meijón M, Satbhai SB, Tsuchimatsu T, Busch W (2014) Genome-wide association study using cellular traits identifies a new regulator of root development in Arabidopsis. Nat Genet 46: 77–81. Available: http://dx.doi.org/10.1038/ng.2824. Accessed 21 July 2014.

28. Slovak R, Göschl C, Su X, Shimotani K, Shiina T, et al. (2014) A Scalable Open-Source Pipeline for Large-Scale Root Phenotyping of Arabidopsis. Plant Cell 26: 2390–2403. Available: http://www.plantcell.org/content/26/6/2390.abstract?etoc. Accessed 22 July 2014.

29. Rosas U, Cibrian-Jaramillo A, Ristova D, Banta JA, Gifford ML, et al. (2013) Integration of responses within and across Arabidopsis natural accessions uncovers loci controlling root systems architecture. Proc Natl Acad Sci U S A 110: 15133–15138. Available: http://www.pnas.org/content/110/37/15133.short. Accessed 26 August 2014.

30. Fujikura U, Horiguchi G, Ponce MR, Micol JL, Tsukaya H (2009) Coordination of cell proliferation and cell expansion mediated by ribosome- related processes in the leaves of Arabidopsis thaliana. Plant J 59: 499–508.

31. Geiler-Samerotte K, Bauer C, Li S, Ziv N, Gresham D, et al. (2013) The details in the distributions: why and how to study phenotypic variability. Curr Opin Biotechnol 24: 752–759. Available: http://dx.doi.org/10.1016/j.copbio.2013.03.010. Accessed 24 August 2013.

32. Ordas B, Malvar RA, Hill WG (2008) Genetic variation and quantitative trait loci associated with developmental stability and the environmental correlation between traits

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted October 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

in maize. Genet Res 90: 385–395. Available: http://www.ncbi.nlm.nih.gov/pubmed/19061529. Accessed 27 May 2014.

33. Rönnegård L, Shen X, Alam M (2010) hglm: A Package for Fitting Hierarchical Generalized Linear Models. R J 2.

34. Aulchenko YS, Ripke S, Isaacs A, Van Duijn CM (2007) GenABEL: an R library for genome-wide association analysis. Bioinformatics 23: 1294–1296.

35. Shen X, Alam M, Fikse F, Rönnegård L (2013) A novel generalized ridge regression method for quantitative genetics. Genetics 193: 1255–1268.

36. Kumar P, Henikoff S, Ng PC (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 4: 1073–1081. Available: http://dx.doi.org/10.1038/nprot.2009.86. Accessed 14 July 2014.

37. Kim S, Plagnol V, Hu TT, Toomajian C, Clark RM, et al. (2007) Recombination and linkage disequilibrium in Arabidopsis thaliana. Nat Genet 39: 1151–1155.

38. Cao J, Schneeberger K, Ossowski S, Günther T, Bender S, et al. (2011) Whole-genome sequencing of multiple Arabidopsis thaliana populations. Nat Genet 43: 956–963.

39. Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, et al. (2012) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res 40: D1202–D1210.

40. Scheible W-R, Pauly M (2004) Glycosyltransferases and cell wall biosynthesis: novel players and insights. Curr Opin Plant Biol 7: 285–295.

41. Husar S, Berthiller F, Fujioka S, Rozhon W, Khan M, et al. (2011) Overexpression of the UGT73C6 alters brassinosteroid glucoside formation in Arabidopsis thaliana. BMC Plant Biol 11: 51. Available: http://www.biomedcentral.com/1471-2229/11/51. Accessed 3 December 2013.

42. Poppenberger B, Fujioka S, Soeno K, George GL, Vaistij FE, et al. (2005) The UGT73C5 of Arabidopsis thaliana glucosylates brassinosteroids. Proc Natl Acad Sci U S A 102: 15253–15258. Available: http://www.pnas.org/content/102/42/15253.full. Accessed 3 December 2013.

43. Priest DM, Jackson RG, Ashford DA, Abrams SR, Bowles DJ (2005) The use of abscisic acid analogues to analyse the substrate selectivity of UGT71B6, a UDP-glycosyltransferase of Arabidopsis thaliana. Available: http://www.sciencedirect.com/science/article/pii/S0014579305008537. Accessed 3 December 2013.

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted October 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

44. Jin S-H, Ma X-M, Han P, Wang B, Sun Y-G, et al. (2013) UGT74D1 is a novel auxin glycosyltransferase from Arabidopsis thaliana. PLoS One 8: e61705. Available: http://dx.plos.org/10.1371/journal.pone.0061705. Accessed 5 December 2013.

45. Kleczkowski K, Schell J, Bandur R (1995) Phytohormone conjugates: nature and function. CRC Crit Rev Plant Sci 14: 283–298.

46. Sullivan A, Arsovski A, Lempe J, Bubb K, Weirauch M, et al. (2014) Mapping and dynamics of regulatory DNA and transcription factor networks in A. thaliana. Cell Rep In press.

47. Winter D, Vinegar B, Nahal H, Ammar R, Wilson G V, et al. (2007) An "Electronic Fluorescent Pictograph" browser for exploring and analyzing large-scale biological data sets. PLoS One 2: e718.

48. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81: 559–575.

49. Ossowski S, Schneeberger K, Clark RM, Lanz C, Warthmann N, et al. (2008) Sequencing of natural strains of Arabidopsis thaliana with short reads. Genome Res 18: 2024–2033.

50. Aeong Oh S, Park J-H, In Lee G, Hee Paek K, Ki Park S, et al. (1997) Identification of three genetic loci controlling leaf senescence in Arabidopsis thaliana. Plant J 12: 527–535. Available: http://dx.doi.org/10.1046/j.1365-313X.1997.00489.x.

51. Jack T (2004) Molecular and genetic mechanisms of floral control. Plant Cell 16 Suppl: S1–17. Available: http://www.plantcell.org/content/16/suppl_1/S1.long. Accessed 3 June 2014.

52. Sohn EJ, Rojas-Pierce M, Pan S, Carter C, Serrano-Mislata A, et al. (2007) The shoot meristem identity gene TFL1 is involved in flower development and trafficking to the protein storage vacuole. Proc Natl Acad Sci U S A 104: 18801–18806. Available: http://www.pnas.org/content/104/47/18801.full. Accessed 4 December 2013.

53. Larsson AS, Landberg K, Meeks-Wagner DR (1998) The TERMINAL FLOWER2 (TFL2) Gene Controls the Reproductive Transition and Meristem Identity in Arabidopsis thaliana. Genetics 149: 597–605. Available: http://www.genetics.org/content/149/2/597.full. Accessed 5 December 2013.

54. Li Y, Baldauf S, Lim E-K, Bowles DJ (2001) Phylogenetic analysis of the UDP-glycosyltransferase multigene family of Arabidopsis thaliana. J Biol Chem 276: 4338–4343.

55. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, et al. (2008) NCBI BLAST: a better web interface. Nucleic Acids Res 36: W5–W9.

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted October 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

56. Masel J, Siegal ML (2009) Robustness: mechanisms and consequences. Trends Genet 25: 395–403.

57. Bolle C, Huep G, Kleinbölting N, Haberer G, Mayer K, et al. (2013) GABI-DUPLO: a collection of double mutants to overcome genetic redundancy in Arabidopsis thaliana. Plant J 75: 157–171. Available: http://www.ncbi.nlm.nih.gov/pubmed/23573814. Accessed 28 August 2014.

58. Im Kim J, Murphy AS, Baek D, Lee S-W, Yun D-J, et al. (2011) YUCCA6 over-expression demonstrates auxin function in delaying leaf senescence in Arabidopsis thaliana. J Exp Bot 62: 3981–3992.

59. Gazzani S, Gendall AR, Lister C, Dean C (2003) Analysis of the molecular basis of flowering time variation in Arabidopsis accessions. Plant Physiol 132: 1107–1114.

60. Wang X, Weigel D, Smith LM (2013) Transposon variants and their effects on gene expression in Arabidopsis. PLoS Genet 9: e1003255. Available: http://dx.plos.org/10.1371/journal.pgen.1003255. Accessed 11 July 2014.

61. Plantegenet S, Weber J, Goldstein DR, Zeller G, Nussbaumer C, et al. (2009) Comprehensive analysis of Arabidopsis expression level polymorphisms with simple inheritance. Mol Syst Biol 5: 242. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2657532&tool=pmcentrez&rendertype=abstract. Accessed 2 September 2014.

62. Zhang X, Cal AJ, Borevitz JO (2011) Genetic architecture of regulatory variation in Arabidopsis thaliana. Genome Res 21: 725–733. Available: http://genome.cshlp.org/content/21/5/725.full.html. Accessed 2 September 2014.

63. Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. Nat Methods 9: 671–675.

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted October 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure Legends

Figure 1. The distributions of within-accession root length mean and variance in a sample of 93 natural *A. thaliana* accessions. There is a large dispersion of the root length variance (A) and mean (B) at day seven among the accessions (*n* = 3 x 70 dark grown seedlings / genotype).

Figure 2. An association on chromosome 3 (17,201,307 bp) identifies At3g46700 as a candidate gene to influence root length developmental noise. A) Using a whole-genome generalized ridge regression model, we identified a significant SNP (3_17201307) associated with root variance. B) The minor allele (G) displays a greater root length variance. C) Accessions carrying a stop codon in At3g46700 tend to have lower root variance (p = 0.06, t-test).

Figure 3. Four pairs of interacting loci are associated with root mean. A) The five *A. thaliana* chromosomes make up the x- and y-axes. The positions of interacting SNPs are indicated by a black dot. Solid lines indicate support for an interaction by more than one linked SNP and dotted lines indicate support by a single SNP. B) Genotype-phenotype (GP) map for the root mean differences among the four genotype combinations from the interaction between a SNP on chromosome 1, 17,257,526 bp and two linked SNPs on chromosome 5 at 15,862,026 and 15,862,525 bp. The major allele is indicated by -1, and the minor allele is indicated by 1. This GP-map is a representative example of what is observed also for the other pairs (Figure S5).

Figure 4. Four candidate genes exhibit reduced root growth in T-DNA insertion lines. T-DNA lines for candidate genes in six of the seven epistatic regions were tested for root length differences to wild-type (Col-0). The gene models for the tested T-DNA insertion lines are

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted October 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

pictured. For four of the six loci, a T-DNA line exhibited significantly different root length from Col-0 (significantly different lines are marked with a * in the figure).

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted October 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1. Estimated narrow (h^2) and broad sense (H^2) heritabilities of root length mean and variance in a population of 93 natural *A. thaliana* accessions. The broad-sense heritability was intermediate for root length mean, with a negligible additive contribution, and very low for root length variance. This suggests a highly epistatic genetic architecture for root-length mean.

Trait	H^2		h^2
	ANOVA[1]	hglm[2]	hglm[2]
Root length mean	0.25 (p = 0)	0.24 (p = 7.2×10^{-03})[3]	0.015 (p = 0.79)[3]
Root length variance	NA	0.0016 (p = 0.23)[3]	0.0011 (p = 0.10)[3]

[1]Estimated using ANOVA with genotype as a fixed effect, [2]Estimated using the R/hglm package (hglm) by fitting a linear mixed model including both additive and epistatic effects as random effects, [3]p-values from Wald tests for the heritability being larger than zero, where the standard errors were estimated via jackknife resampling.

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted October 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 2. Exhaustive two-dimensional GWAS-scan for epistasis identifies six significant interactions, representing four unique pairs, associated with the root length mean.

Pair[1]	SNP1[1]	SNP2[1]	Interaction effect (mm)	p-value[2]
1	rs346846662 (3_66596)	rs347034954 (3_9272294)/ rs347048583 (3_9273674)	0.55	1.5×10^{-10}
2	rs347005428 (1_17257526)	rs346595296 (5_15862026)/ rs346511785 (5_15862525)	0.63	1.7×10^{-10}
3	rs347374065 (3_10891195)	rs347377866 (5_1027939)	0.63	2.3×10^{-10}
4	rs346846662 (3_66596)	rs347177728 (5_18241640)	0.63	3.0×10^{-10}

[1]Name of associated SNP(s) at the epistatic locus as on the Affymetrix single nucleotide polymorphism (SNP) chip (our abbreviation as Chromosome_PositionInBp), [2]p-value for the interaction effect of the two loci [48].

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted October 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 3. Summary of identified polymorphisms in LD with the leading SNPs from the epistatic GWAS analysis. Four unique interacting pairs of loci were significantly associated with root length (Table 1). Here, the genes located within 1000bp of the leading SNPs, together with the mutational effects of the identified polymorphisms, are listed.

Chr	SNP positions (bp)	In pair	Locus	# Mutations (# in LD)[1]	Mutation Effects
3	66596	1, 4	At3g01180	1 (0)	1 upstream
			At3g01185	1 (0)	1 syn
3	9,272,294 9,273,674	1	At3g25520	4 (21)	1 intron, 3 upstream
			At3g25530	14 (21)	1 mis, 1 splice, 1 syn, 2 intron, 9 upstream
			At3g25540	17 (21)	1 mis, 4 syn, 2 intron, 5 5'UTR, 5 upstream
			At3g25545	3 (21)	3 upstream
5	18,241,640	4	At5g45120	6 (5)	2 mis, 4 syn
1	17,257,526	2	At1g46624	1 (0)	1 non-coding exon
5	15,862,026 15,862,525	2	At5g39610	2 (4)	6 upstream[3]
			At5g39620	2 (4)	6 downstream[3]
3	10,891,195	3	At3g28865	4 (7)	4 non-coding exon
			At3g28870	2 (7)	1 mis, 1 syn
			At3g28880	3 (7)	3 upstream
5	1,027,939	3	At5g03840	1 (1)	1 upstream
			At5g03850	1 (1)	1 downstream

[1]The number of polymorphisms associated with each listed gene, with the number of these polymorphisms that are in LD with the leading SNP in parenthesis. Some mutations are associated with more than one gene. [2]Abbreviations for mutational effects: syn = synonymous variant; mis = missense variant; splice = splice site variant. [3]SNPs more than 1000bp away from the nearest gene

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted October 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

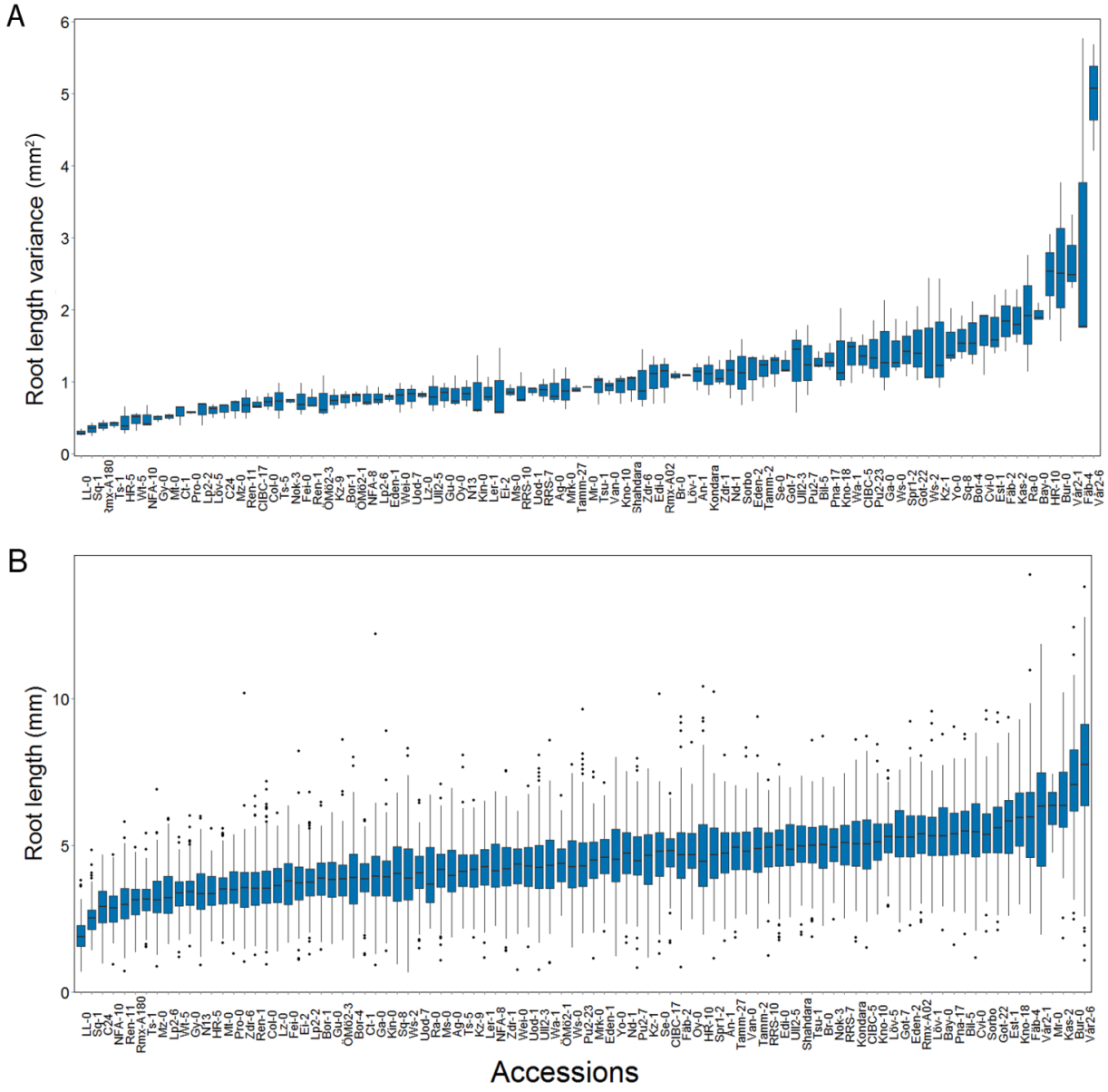
Table 4. Function and expression patterns of genes in LD with leading SNP from epistatic GWAS analysis. The gene names or proposed function is listed for all genes harboring polymorphisms in LD with the leading epistatic SNPs in the whole genome interaction analysis.

Chr	Position (bp)	In pair	Locus	Gene/Putative function	Highest expression tissue[1]
3	66596	1, 4	At3g01180	*SS2*	cauline leaves
			At3g01185	unknown	NA
3	9,272,294/	1	At3g25520	*ATL5*	apical meristem, seedling root
	9,273,674		At3g25530	*LOH1*	rosette leaves
			At3g25540	*GR1*	seedling root, imbibed seed
			At3g25545	unknown	seedling root (vasculature), imbibed seed
5	18,241,640	4	At5g45120	eukaryotic aspartyl protease family protein	adult root (meristem), imbibed seed
1	17,257,526	2	At1g46624	gypsy-like retrotransposon	NA
5	15,862,026/	2	At5g39610	*NAC6*	seedling root (vasculature)
	15,862,525		At5g39620	RAB GTPase homolog G1	adult root vasculature, mature seed
3	10,891,195	3	At3g28865	LINE retrotransposon	NA
			At3g28870	histone deacetylase	mature seed
			At3g28880	Ankyrin repeat family protein	NA
5	1,027,939	3	At5g03840	*TFL1*	adult root vasculature (mature zone)
			At5g03850	nucleic acid-binding, OB-fold-like protein	seedling root, apical meristem

[1] Locations of high expression were obtained from the BAR eFP *Arabidopsis* Browser [47].

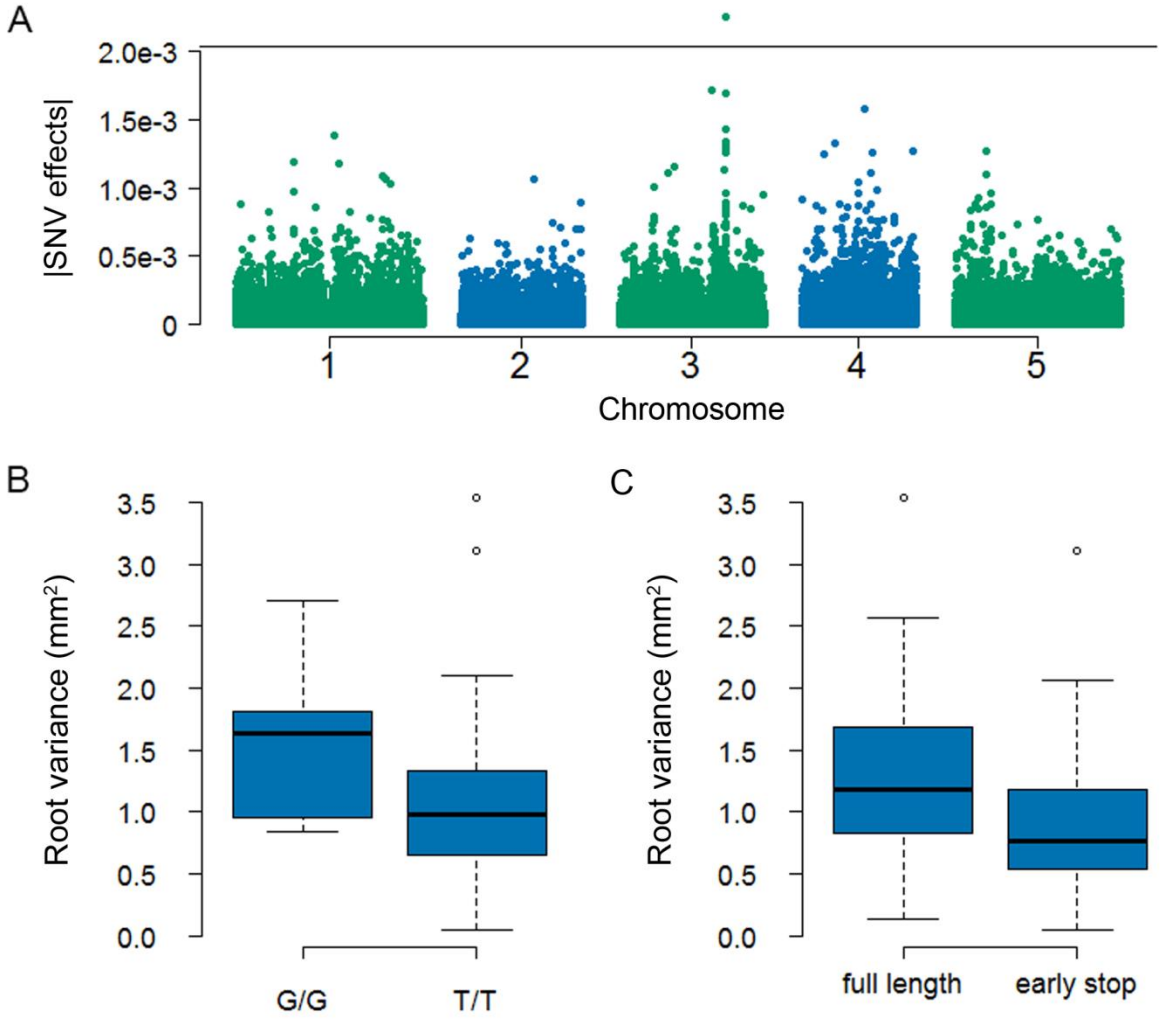
bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted October 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 1



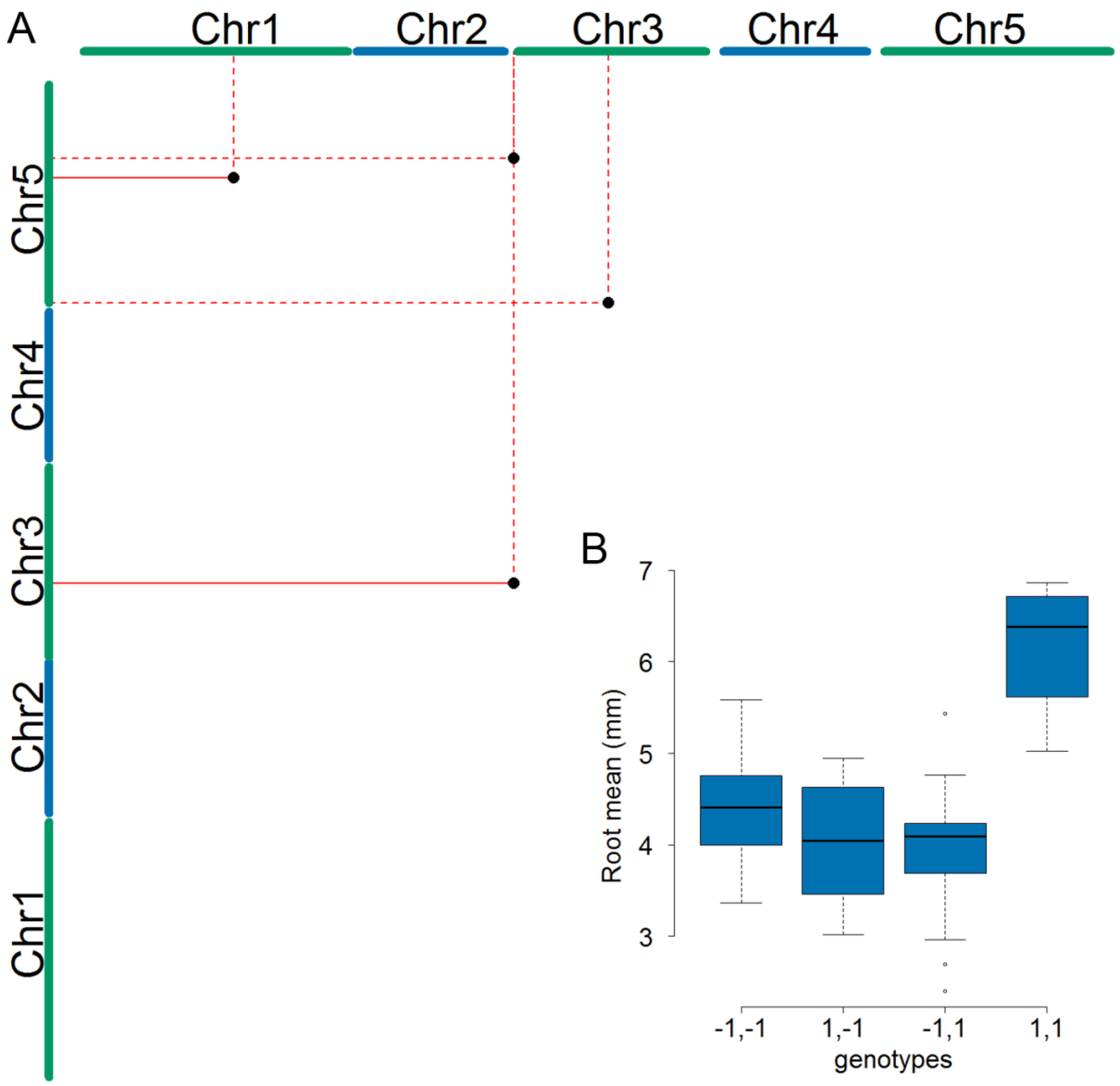
bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted October 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 2



bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted October 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 3



bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted October 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

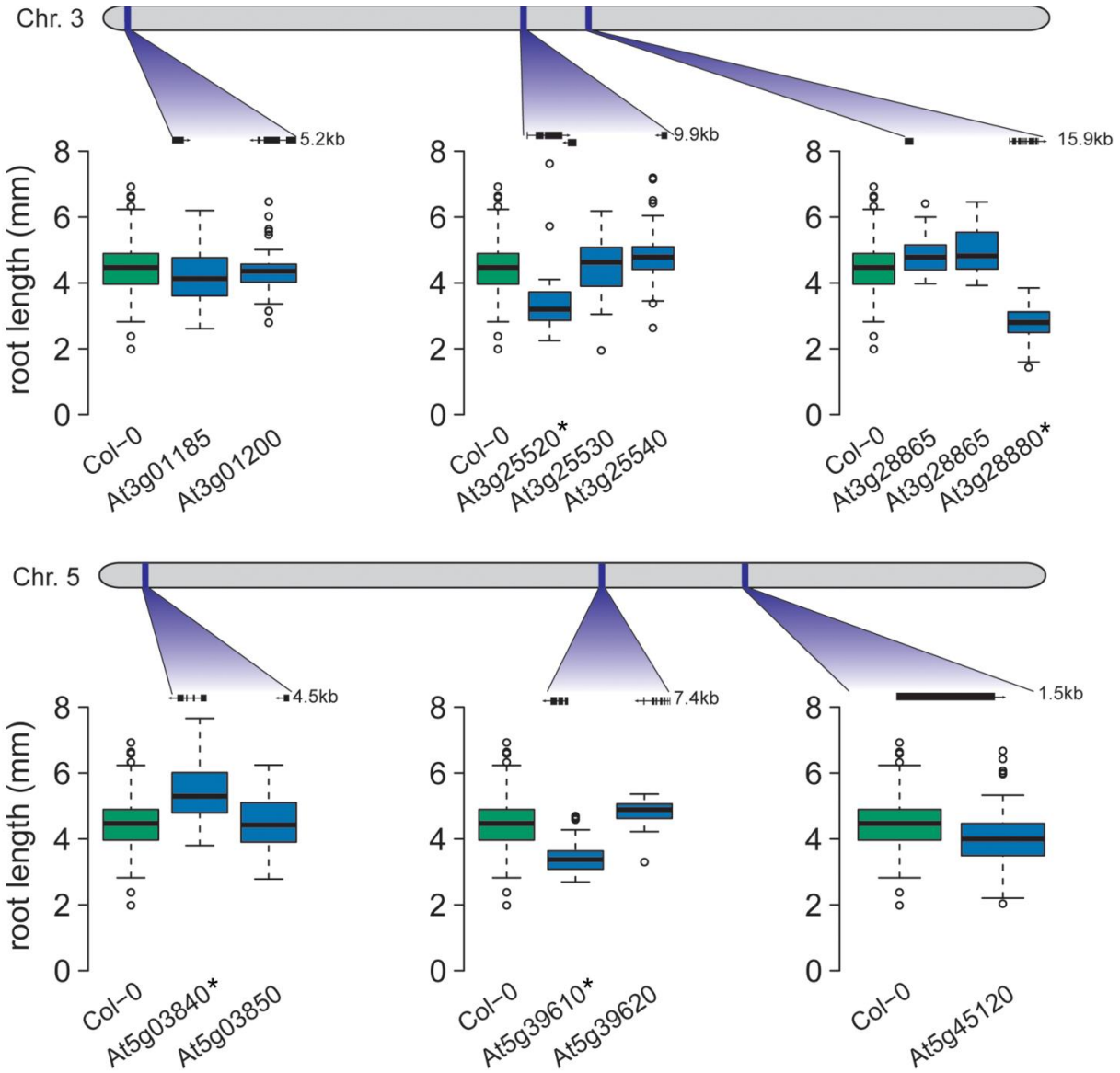



Figure 4