bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted September 4, 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 The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.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

Highly epistatic genetic architecture of root length in *Arabidopsis thaliana*

Authors

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

Affiliations

[1]Department of Genome Sciences, University of Washington

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

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

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 capturing epistatic interactions, overcome these challenges. Specifically, we examined the genetic determinants of *Arabidopsis thaliana* root length, considering both trait mean and variance. Analysis of narrow-and broad-sense heritability of mean root length identified a large contribution of non-additive variation and a low contribution of additive variation. Also, no loci were found to contribute to mean root length using a standard additive model based genome-

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted September 4, 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 The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.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.

wide association analysis (GWAS). We could, however, identify one locus regulating developmental noise and seven loci contributing to root-length through epistatic interactions, and four of these were also experimentally confirmed. The candidate 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. The mutant analysis of candidate genes within the seven epistatic loci implicated four genes in 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 *A. thaliana*.

Author summary

Complex traits, such as height and many human diseases, often arise through the action and interaction of many genes and environmental factors. Classic genetic approaches to identify the contributing genes assume that these factors act additively. Recent methods such as genome-wide association studies also rely on this assumption. However, additive models of complex traits do not reflect that genes can interact and affect each other's activity in a non-additive manner. In this study, we use the model plant *Arabidopsis thaliana* to determine the additive and non-additive contributions to the complex trait root length. Surprisingly, much of the observed phenotypic variation in root length across genetically divergent strains can be explained by genetic interactions, and here we validate four epistatic genes using mutant analysis. Three of these genes are newly implicated in root development. Taken together, our results emphasize the importance of considering both non-additive and additive effects when dissecting complex traits.

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted September 4, 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.

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]. Epistatic interactions between loci are pervasive in human and model organism complex traits whenever attempts are made to identify them [3]. To obtain a more complete compilation of the loci contributing to the variation of a complex trait, one needs to consider developmental noise, epistatic interactions and additive genetics.

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

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted September 4, 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 levels have been primarily identified with mutant screens, candidate gene approaches [2,7,12,13], and QTL mapping [8–10,14–16]. Similarly, efforts to identify epistatic loci are often targeted, considering only candidate loci or pathways [3,17–22]. 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 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 both due to 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, especially in inbreeding organisms with smaller genomes in which the number of multi-locus genotype-classes is smaller, phenotypes can be repeatedly measured for each genotype, and for which the multiple-testing penalty is lower. *A. thaliana* has previously been used for standard GWAS [25–28]. As *A. thaliana* has a small genome (~125 Mb), large numbers of readily available inbred accessions, and a large knowledge base on how to efficiently score quantitative

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted September 4, 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.

phenotypes in large numbers of individuals, it is a particularly well-suited model for performing GWAS analyses of developmental noise and exhaustive scans for epistasis.

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 consistent with a low narrow-sense 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. Notably, the broad-sense heritability of *A. thaliana* mean root length was considerably larger than the narrow-sense heritability, 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 assumption 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 candidate gene known to affect root length [29]. 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. 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.

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted September 4, 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.

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 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, but not trait variances [7,8,25]. The root length variance 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. We grew seedlings in 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 [7,30]. We examined the reproducibility of variance 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 length variances across accessions varied significantly ($p = 7.5*10^{-239}$, Brown-Forsythe test), suggests that root length variance 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 variances with even greater accuracy. We found a strong correlation between the mean variance of the three biological replicates and the pooled variance for each

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted September 4, 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.

accession ($p = 1.8*10^{-68}$, $R^2 = 0.97$, Figure S2a) and proceeded to use pooled variance measurements in all further analyses.

Several previous studies of loci associated with developmental noise found that many of these also affect trait means [8,10,31]. This potentially confounding relationship between trait mean and variance is a concern in examining developmental noise [30] as classical statistical models operate on the assumption that mean and variance are independent. If this assumption is violated, other models must be used. We confirmed that mean and variance of root length were independent in our data set by determining that mean root length and variance were not correlated ($p = 0.11$, $R^2 = 0.02$, Figure S2b). Mean root length varied significantly across accessions ($p<2.2*10^{-16}$, ANOVA, Figure 1b). The genetic architectures underlying mean root length and variance 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 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 [32]. The broad sense heritability was 0.23 for mean root length and 0.00159 for root length variance (Table 1). We further partitioned the genetic variance explained into additive and epistatic variance. For mean root length, a larger fraction of the variance was explained through epistatic effects (0.22) than by additive effects (0.01). For root length variance, a larger fraction of the genetic variation was additive (0.00109) than epistatic (0.00050). These heritability patterns suggested that GWAS analysis would be unlikely to uncover loci explaining major portions of phenotypic variation in root length mean

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted September 4, 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.

and variance (*i.e.* 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 [33]. 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 [34]. 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 narrow-sense heritability of 0.01 for mean root length, 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 GWAS models described above. 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 ($p = 1.61*10^{-7}$, threshold = $2.336*10^{-7}$). Using the HEM GWAS model, however, the minor allele for this SNP on chromosome 3 at position 17,201,307bp (3_17201307) was significantly

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted September 4, 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.

associated with greater developmental noise in root length (|SNP effect| = $2.26*10^{-3}$, threshold = $2.04*10^{-3}$, Figure 2a, b).

The SNP 3_17201307 encodes a missense variant in the gene At3g46700. We explored whether this associated SNP was in linkage disequilibrium (LD) with any other SNPs ± 15 kb away - the LD typically decays within 10kb [35] in *A. thaliana*. It was in LD ($R^2 > 0.8$) with 10 other SNPs identified in the 1001 Genomes Project within this 30kb 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 [36]. UDP-glycosyltransferases catalyze the enzymatic addition of sugars onto a variety of other molecules, such as secreted components of the plant cell wall [37] and phytohormones [38–42]. 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 [43]. The last linked SNP lies in the coding region of At3g46690, causing a S->T missense change, predicted to be tolerated [44]. 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 [45].

Further explorations of the natural variation present in At3g46700 and At3g46690 among the 93 accessions using data from the 1001 Genomes Project [46] revealed the presence of two stop 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 the root length variance

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted September 4, 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.

among accessions containing a stop codon to the accessions without stop codons. Root length variance was indeed greater among accessions with a functional At3g46700, but not significantly so (Figure 2c, p=0.06, t-test, n=63). As this accession set was small, further analysis is needed to fully explore whether functional At3g46700 decreases root length variance.

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) [47]. 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 the 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 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 [35]. In total, the epistatic analysis included approximately 78 million independent tests. Six SNP-pairs were significantly associated with mean root length after accounting for multiple testing (Figure 3a, Table 2, Figure S5). These pairs represented four unique combinations of loci,

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted September 4, 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.

and seven genomic locations; two of these pairs involved 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 variance in mean root length, which is most of the estimated epistatic contribution to the genetic variance in mean root length (Table 1). We further examined mean root lengths and genotype frequencies for each two-locus genotype of the four significant SNP pairs (Figure 3b) and observed the same genotype-phenotype 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 fine-map these significant epistatic associations, we added 13 million SNPs obtained by whole-genome re-sequencing in the 1001 Genomes effort to our analysis [48]. SNPs that were in LD ($R^2 > 0.8$) and located within 5kb upstream or downstream of the leading SNP were considered further (Table 2). For each pair, the 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

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted September 4, 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.

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 decrease root length

Two linked SNPs on chromosome 3 (3_9272294 and 3_9273674) 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 exhibited significantly shorter roots than wild type (Tukey's post-hoc test); the insertion mutants for At3g25530 and At3g25540 did not differ from wild type (Figure 4).

At3g25520 encodes ATL5/OLI5/PGY3, a 5S rRNA binding protein whose promoter is highly accessible in seedling roots [43], corresponding to expression in roots [45]. Our finding agrees with earlier studies, in which the loss-of-function mutant *oli5-1* exhibits shorter primary roots than wild-type [29]. 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 [45]. In the tested accessions, this gene contains three SNPs in its putative promoter, motivating future studies to explore its role in root growth.

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted September 4, 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 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) interacted with an SNP on chromosome 1 (17,257,526 bp - SNP 1_17257526). Both of these linked SNPs, as well as the other SNPs in high LD, were intergenic with the nearest gene being *NAC6* (At5g39610; Table 4), which is located 1.9 kb away. *NAC6* is a transcription factor regulating leaf senescence [49] and is highly expressed in senescing leaves and in maturing seeds [45]. We phenotyped the available At5g39610 insertion mutant and found a significant decrease in mean root length (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 [43]. We speculate that the polymorphisms in our tested accessions alter the regulatory potential of these regions.

A functional role for the previously uncharacterized gene At3g28880 in root length

A third locus on chromosome 3 (10,891,195 - SNP 3_10891195) interacted with a locus on chromosome 5 (1,027,939 bp - SNP 5_1027939). SNPs in LD with 3_10891195 were found in three genes: At3g28865, At3g28870, and At3g28880 (Table 4). None of these genes had been previously 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

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted September 4, 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.

decreased mean root length compared to wild-type controls (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 [43]. 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 [36]. 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.

A novel function for TFL1 with its mutant increasing mean root length

The SNP 5_1027939 was detected via its epistatic interaction with the At3g28880 locus on chromosome 3. This polymorphism and the only other polymorphism in LD with it, are both intergenic between the genes At5g03840 and At5g03850 (Table 4). Phenotyping the insertion mutants for these positional candidate genes revealed that the mutant for At5g03840, which encodes TFL1, showed significantly longer roots than wild-type controls (Figure 4), whereas no significant phenotypic effect was found for the At5g03850 mutant.

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 [50]. It also plays a role in regulating protein storage in vegetative tissues, such as roots or seeds [51]. *TFL1* is highly expressed in both adult and seedling roots [45]. Our finding that *TFL1* controls root length, and

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted September 4, 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.

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 [52]. 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 [45]. 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 mean root length and variance. We found that the genetic architecture of mean root length differed greatly from that of root length variance in our analysis population. The narrow-sense heritability of mean root length was about a magnitude higher than for the root length variance (Table 1), consistent with previous studies of trait means and variance [8–10]. 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 readily identified a locus which was significantly associated with root

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted September 4, 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.

developmental noise in spite of this trait's much lower narrow-sense heritability. By taking potential epistatic interactions in shaping mean root length in 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. Consequently, in this study we were able to attribute most of the epistatic genetic variance in a complex trait to the effects of individual loci, whereas the loci underlying the additive genetic variance in mean root length remain unknown. 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.

No individual loci could be mapped for mean root length using an additive model-based GWAS analysis. The reasons for this result are unknown, but a potential explanation is the inability to evaluate rare SNPs due to the small number of considered accessions (93). Alternatively, the additive variance of mean root length may be due to the effects of many loci, each of them contributing additive effects on root length that are too small to be detectable using a stringent Bonferroni threshold.

GWAS have earlier been used to identify loci affecting developmental noise in gene expression [1,8,15]. 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 [53]. Several potential substrates of At3g46700 and At3g46690 are known to be important for root growth, including cell wall components and phytohormones [37–42]. The 25 kb region

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted September 4, 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.

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

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 [55]. Our findings challenge the generality of this assumption, as both the minor allele and 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.

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted September 4, 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.

Functional dissection of epistasis is a daunting task in natural populations. A first step in taking on this task 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 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 [2,7,56]. 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 [57]. We found that natural variation in the *NAC6* promoter was associated with root length through 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 [58]. Further, a recent study found that the presence of polymorphisms in transposons is positively correlated with gene expression variation in adjacent genes [59]. The transposon-rich region is located on a different

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted September 4, 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.

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 expression variation in shaping phenotypic variation [25,60,61]. 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, nor have these genes been previously 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 highly epistatic, 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 a large

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted September 4, 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.

portion of the 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 for root length variance, the more complex multi-marker HEM analysis was more powerful for 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.

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

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted September 4, 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 was measured manually using the 'freehand' function in ImageJ1.46r [62]. Non-germination, missing organs, and delayed development were noted.

Data standardization

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 predictions

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 [32], i.e.

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

where

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

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

$$b \sim N(G \circ G\sigma_a^2)$$

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 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.

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted September 4, 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.

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 associations. For the standard additive GWAS with population structure controlled, we used the function egscore from R/GenABEL [33]. The functions bigRR and bigRR_update in the R/bigRR package [34] 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 [47] 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 4 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 and 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-

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted September 4, 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.

testing corrected significance threshold for this epistatic analysis ($p = 3.2*10^{-10}$) by estimating the number of independent tests based on the number of estimated LD blocks in the genome, where the *A. thaliana* genome is approximately 125Mb with average LD-blocks of about 10kb (12,500; [35]), and applying a Bonferroni correction for the 78 million independent tests performed. Using data on the sixty-three accessions from the GWAS that were unambiguously identified in the 1001 Genomes data [46], we then identified additional SNPs in LD with the leading SNP using the function *LD* in the *genetics* package in R across a region of 10kb 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 [45].

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 across the T-DNA lines and wild-type (Col-0).

Acknowledgments

We thank Mats Pettersson, Ronnie Nelson, Karla Schultz, and members of the Queitsch lab. We recognize funding from the National Science Foundation Graduation Research Fellowship (DGE-1256082) (J.L.), the National Institutes of Health New Innovator Award 13

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted September 4, 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.

(DP2OD008371) (C.Q), The Swedish Foundation for Strategic Research (Future Research Leader Grant to Ö.C.), and FORMAS (Ö.C.).

Figure Legends

Figure 1. Across accessions, there is a broad distribution in root length and variance. A) Root length variance was measured in three replicates of 70 individuals. B) Root length was measured in ~210 dark grown seedlings at seven days.

Figure 2. A variant on chromosome 3 (3_17201307) is associated with root variance. A) Using a whole-genome generalized ridge regression model in which SNPs were random effects, a significant SNP is detected in association with root variance. B) The minor allele (G) is found associated with greater root length variance. C) Accessions containing a stop codon in At3g46700 tend to have lower root variance ($p = 0.06$, t-test).

Figure 3. Four distinct genetic interactions 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 point. Solid lines indicate support for an interaction by more than one linked SNP and dotted lines indicate support by a single SNP. B) Pictured is a representative example of the root mean differences among the four genotype combinations from the interaction between 1_17257526 and 5_15862026 /5_15862525. The major allele is indicated by -1, and the minor allele is indicated by 1.

Figure 4. Candidate genes exhibit reduced root growth. T-DNA lines for several candidate genes on chromosomes 3 and 5 have been tested for root length in comparison to wild-type (Col-

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted September 4, 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.

0). The gene models for the tested T-DNA insertion lines are pictured. For four of the six loci, a T-DNA line exhibited significantly different root length from Col-0.

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted September 4, 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.

	hglm			ANOVA
root length	total	additive	epistatic	H^2
Mean	0.23	0.01	0.22	0.25
Variance	0.00159	0.00109	0.00050	NA

Table 1. Heritability of root length mean and root length variance. Heritability was estimated using a linear mixed model including both additive and epistatic effects as random effects using the R/hglm package. Root length mean is heritable with a large epistatic contribution, whereas root length variance has a low heritability. A similar H^2 was estimated for root length mean using ANOVA with genotype as a fixed effect.

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted September 4, 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.

Pair	SNP1	SNP2	Interaction effect size (mm)	p-value
1	3_66596	3_9272294	0.5548	$1.48*10^{-10}$
1	3_66596	3_9273674	0.5548	$1.48*10^{-10}$
2	1_17257526	5_15862026	0.6337	$1.67*10^{-10}$
2	1_17257526	5_15862525	0.6337	$1.67*10^{-10}$
3	3_10891195	5_1027939	0.6256	$2.25*10^{-10}$
4	3_66596	5_18241640	0.6251	$3.03*10^{-10}$

Table 2. Six significant interactions were associated with root length mean.

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted September 4, 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.

Pair	Chr	Position (bp)	# linked mutations	Locus	#Mut	Mutation Effects
1,4	3	66596	0	At3g01180	1	1 upstream
				At3g01185	1	1 syn
1	3	9272294/9273674	21	At3g25520	4	1 intron, 3 upstream
				At3g25530	14	1 mis, 1 splice, 1 syn, 2 intron, 9 upstream
				At3g25540	17	1 mis, 4 syn, 2 intron, 5 5'UTR, 5 upstream
				At3g25545	3	3 upstream
4	5	18241640	5	At5g45120	6	2 mis, 4 syn
2	1	17257526	0	At1g46624	1	1 non-coding exon
2	5	15862026/15862525	4	At5g39610	1	1 upstream*
				At5g39620	1	1 downstream*
3	3	10891195	7	At3g28865	4	4 non-coding exon
				At3g28870	2	1 mis, 1 syn
				At3g28880	3	3 upstream
3	5	1027939	1	At5g03840	1	1 upstream
				At5g03850	1	1 downstream

Table 3. Variants in linkage with leading interacting SNPs from whole genome interaction analysis. Six interacting pairs of loci were found significantly associated with root length, as displayed in Table 1. The genes within 1000bp of the linked SNPs are listed and the mutation mutational effects are labeled: syn = synonymous variant, mis = missense variant, splice = splice site variant. #Mut refers to the number of mutations associated with each gene, and some mutations are associated with more than one gene. *SNPs more than 1000bp away from the nearest gene.

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted September 4, 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.

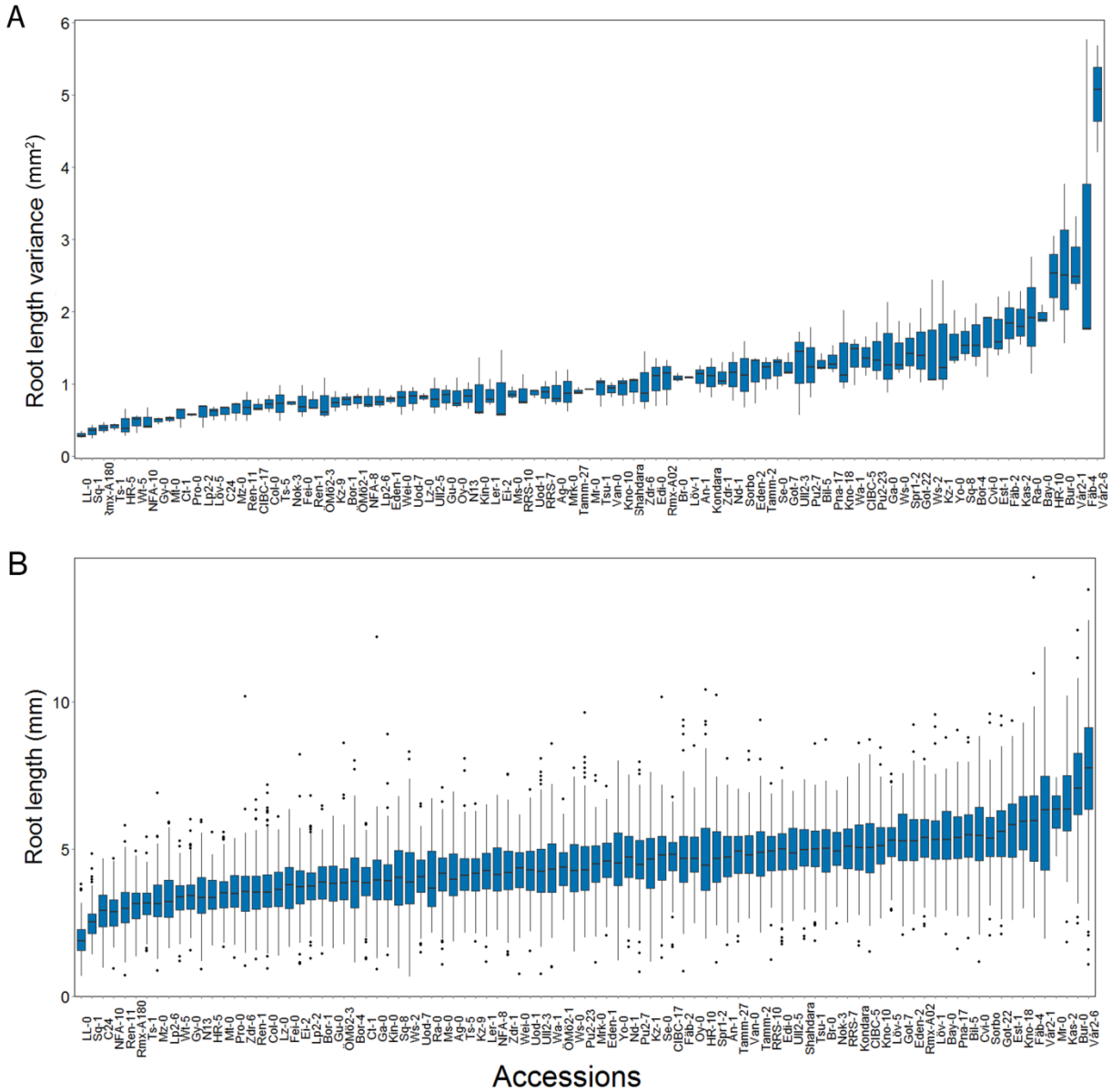
Pair	Chr	Position (bp)	Locus	Gene/Putative function	Highest expression tissues
1,4	3	66596	At3g01180	*SS2*	cauline leaves
			At3g01185	unknown	NA
1	3	9272294/ 9273674	At3g25520	*ATL5*	apical meristem, seedling root
			At3g25530	*LOH1*	rosette leaves
			At3g25540	*GR1*	seedling root, imbibed seed
			At3g25545	unknown	seedling root (vasculature), imbibed seed
4	5	18241640	At5g45120	Eukaryotic aspartyl protease family protein	adult root (meristem), imbibed seed
2	1	17257526	At1g46624	gypsy-like retrotransposon	NA
2	5	15862026/ 15862525	At5g39610	*NAC6*	seedling root (vasculature)
			At5g39620	RAB GTPase homolog G1	adult root vasculature, mature seed
3	3	10891195	At3g28865	LINE retrotransposon	NA
			At3g28870	Histone deacetylase interacting domain	mature seed
			At3g28880	Ankyrin repeat family protein	NA
3	5	1027939	At5g03840	*TFL1*	adult root vasculature (mature zone)
			At5g03850	Nucleic acid-binding, OB-fold-like protein	seedling root, apical meristem

Table 4. Function and expression patterns of genes in linkage with leading SNP from whole genome interaction analysis. The gene names or proposed function is listed for all genes harboring variants in linkage with the leading SNPs found to interact in the whole genome interaction analysis. Locations of high expression were obtained from the BAR eFP *Arabidopsis* Browser [45].

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted September 4, 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.

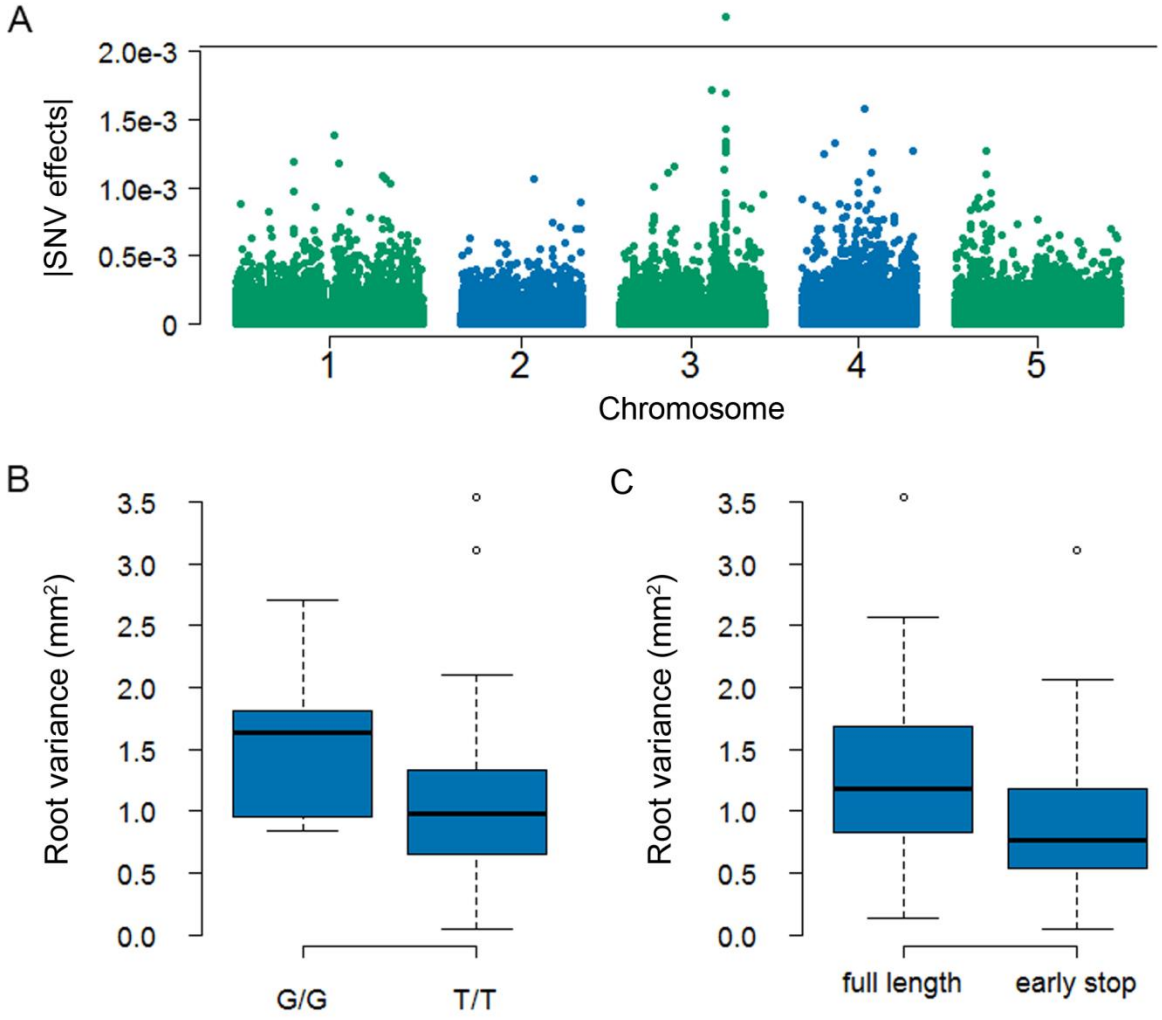
Main Figures

Figure 1



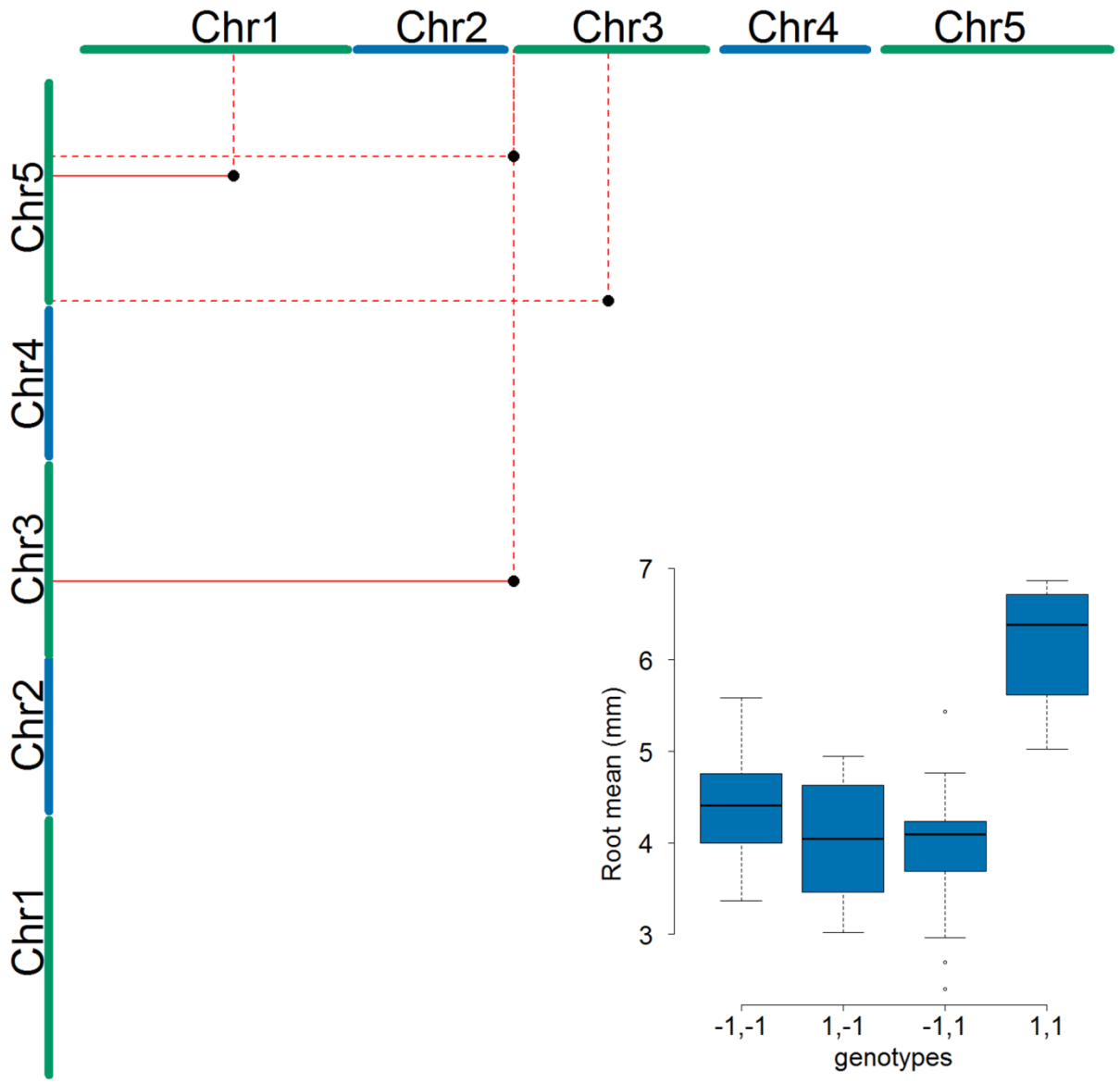
bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted September 4, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 2



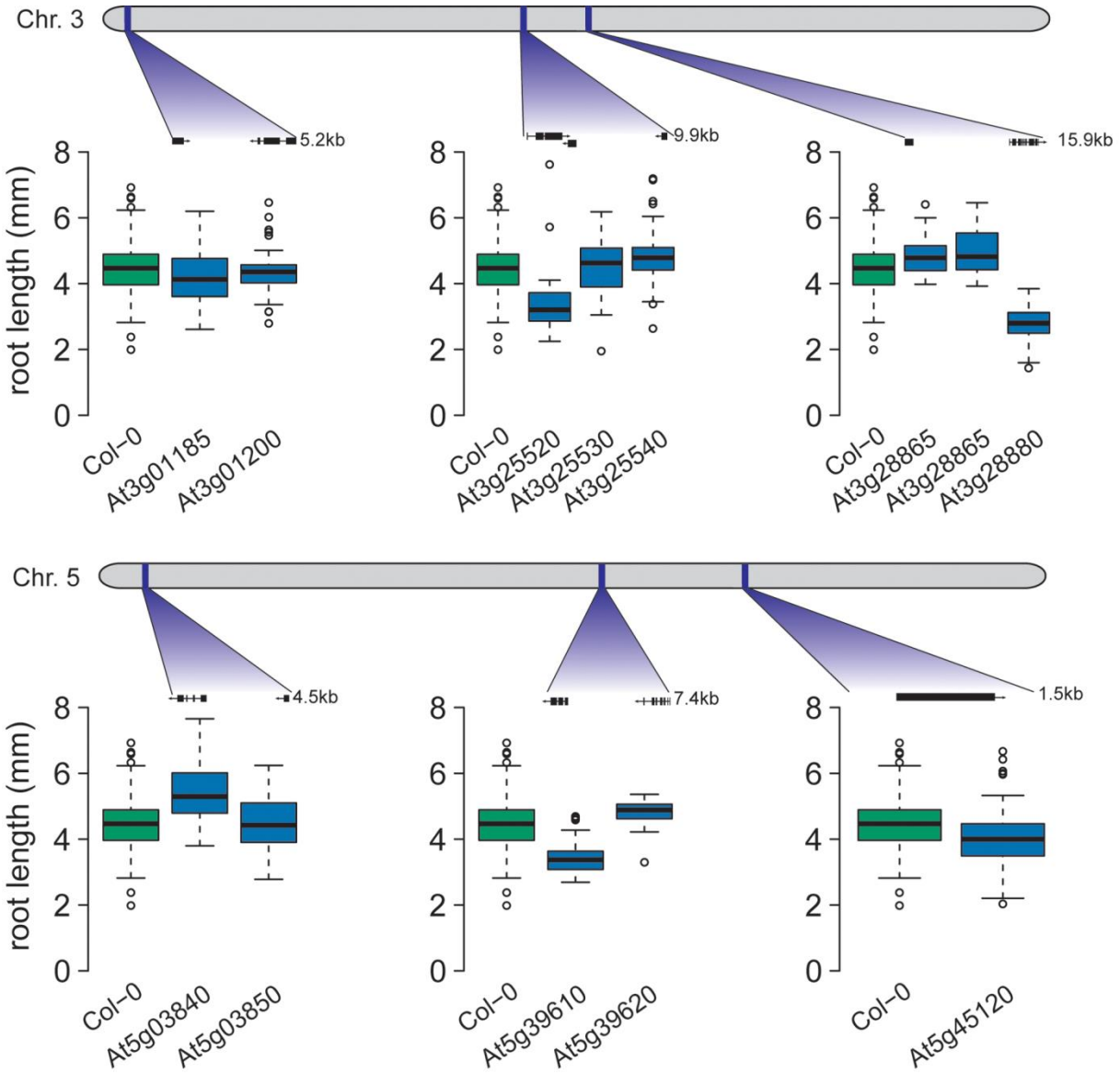
bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted September 4, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 3



bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted September 4, 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



bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted September 4, 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 September 4, 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. 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.

15. 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.

16. 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.

17. 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.

18. 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.

19. 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).

20. 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.

21. 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.

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted September 4, 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. 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.

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. 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.

27. 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.

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. 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.

30. 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.

31. Ordas B, Malvar RA, Hill WG (2008) Genetic variation and quantitative trait loci associated with developmental stability and the environmental correlation between traits in maize. Genet Res 90: 385–395. Available: http://www.ncbi.nlm.nih.gov/pubmed/19061529. Accessed 27 May 2014.

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted September 4, 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.

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

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

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

35. 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.

36. 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.

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

38. 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.

39. 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.

40. 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.

41. 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.

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

43. 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.

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted September 4, 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. 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.

45. 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.

46. 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.

47. 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.

48. 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.

49. 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.

50. 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.

51. 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.

52. 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.

53. 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.

54. 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.

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

bioRxiv preprint doi: https://doi.org/10.1101/008789; this version posted September 4, 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. 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.

57. 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.

58. 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.

59. 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.

60. 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.

61. 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.

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