the plant journal

SEB Society for Experimental Biology

The Plant Journal (2017) **89,** 1225–1235 doi: 10.1111/tpj.13457

RESOURCE

AraQTL – workbench and archive for systems genetics in *Arabidopsis thaliana*

Harm Nijveen[1,2,*], Wilco Ligterink[2], Joost J. B. Keurentjes[3], Olivier Loudet[4], Jiao Long[1,5], Mark G. Sterken[5], Pjotr Prins[6], Henk W. Hilhorst[2], Dick de Ridder[1], Jan E. Kammenga[5] and Basten L. Snoek[5,*]

[1]*Bioinformatics Group, Wageningen University, Droevendaalsesteeg 1, Wageningen NL-6708 PB, The Netherlands,*
[2]*Wageningen Seed Lab, Laboratory of Plant Physiology, Wageningen University, Droevendaalsesteeg 1, Wageningen NL-6708 PB, The Netherlands,*
[3]*Laboratory of Genetics, Wageningen University, Droevendaalsesteeg 1, Wageningen NL-6708 PB, The Netherlands,*
[4]*Institut Jean-Pierre Bourgin, INRA, AgroParisTech, CNRS, Université Paris-Saclay, Versailles 78000, France,*
[5]*Laboratory of Nematology, Wageningen University, Droevendaalsesteeg 1, Wageningen NL-6708 PB, The Netherlands, and*
[6]*Department of Genetics, Genomics, and Informatics, University of Tennessee Health Science Center, Memphis, TN 38163, USA*

Received 7 October 2016; revised 24 November 2016; accepted 6 December 2016; published online 20 December 2016.
*For correspondence (e-mails harm.nijveen@wur.nl; basten.snoek@wur.nl).

SUMMARY

Genetical genomics studies uncover genome-wide genetic interactions between genes and their transcriptional regulators. High-throughput measurement of gene expression in recombinant inbred line populations has enabled investigation of the genetic architecture of variation in gene expression. This has the potential to enrich our understanding of the molecular mechanisms affected by and underlying natural variation. Moreover, it contributes to the systems biology of natural variation, as a substantial number of experiments have resulted in a valuable amount of interconnectable phenotypic, molecular and genotypic data. A number of genetical genomics studies have been published for *Arabidopsis thaliana*, uncovering many expression quantitative trait loci (eQTLs). However, these complex data are not easily accessible to the plant research community, leaving most of the valuable genetic interactions unexplored as cross-analysis of these studies is a major effort. We address this problem with AraQTL (http://www.bioinformatics.nl/AraQTL/), an easily accessible workbench and database for comparative analysis and meta-analysis of all published Arabidopsis eQTL datasets. AraQTL provides a workbench for comparing, re-using and extending upon the results of these experiments. For example, one can easily screen a physical region for specific local eQTLs that could harbour candidate genes for phenotypic QTLs, or detect gene-by-environment interactions by comparing eQTLs under different conditions.

Keywords: genetical genomics, expression quantitative trait loci, system genetics, gene expression, co-expression, transcriptional regulation, *Arabidopsis thaliana*.

INTRODUCTION

Over the last 30 years *Arabidopsis thaliana* has become the primary model plant for quantitative genetics (Koornneef and Meinke, 2010; Weigel, 2012). Many different studies have generated and accumulated a large amount of phenotypic, genotypic and high-throughput molecular data across hundreds of different strains and ecotypes (Alonso-Blanco and Koornneef, 2000; Keurentjes *et al.*, 2007, 2008b; Cao *et al.*, 2011; Gomaa *et al.*, 2011; Kerwin *et al.*, 2011, 2015; Brennan *et al.*, 2014; Joseph *et al.*, 2014; Molenaar and Keurentjes, 2014; Wijnen and Keurentjes, 2014; Taylor-Teeples *et al.*, 2015). This detailed knowledge about its molecular, cellular and physiological processes and phenotypes makes Arabidopsis an ideal model system for studying the effects and causes of natural variation at the system level (Keurentjes *et al.*, 2008a,b, 2011; Alonso-Blanco *et al.*, 2009; Joosen *et al.*, 2009; Keurentjes, 2009;

Stitt *et al.*, 2010; Terpstra *et al.*, 2010; Chan *et al.*, 2011; Kerwin *et al.*, 2011; Trontin *et al.*, 2011; Joseph *et al.*, 2014; Taylor-Teeples *et al.*, 2015). A powerful experimental approach to this makes use of recombinant inbred lines (RILs) derived from two parents to find quantitative trait loci (QTLs) (Jansen and Nap, 2001; Jansen, 2003).

Expression quantitative trait loci

Quantitative trait locus studies use natural genetic variation to identify loci that underlie variation in a quantitative trait between genotypes. At first, traits like fruit weight or flowering time were used. Now, with the availability of microarrays and RNA sequencing (RNA-seq), gene expression can also be used for investigation of QTLs on a genome-wide scale (Jansen and Nap, 2001; Jansen, 2003; Rockman and Kruglyak, 2006). The resulting eQTL (expression QTL) is a polymorphic locus associated with variation in expression of one or more genes. Each gene has its own eQTL profile, depending on the presence of genetic variation in regulatory genes or elements. A polymorphic site in a gene's own promoter that affects that gene's expression can lead to an eQTL at the location of the gene itself, a *cis*-eQTL. A polymorphism in one of the gene's transcriptional regulators that affects this regulator's activity can give rise to an eQTL at the location of the regulator, a *trans*-eQTL. In principle, these eQTLs can be used for the construction of a gene regulatory network (Rockman and Kruglyak, 2006; Keurentjes *et al.*, 2007; Zhu *et al.*, 2008; Terpstra *et al.*, 2010; Vignes *et al.*, 2011) connecting each gene to its regulators. However, this application is limited by the low resolution of the eQTL analysis that typically yields eQTLs spanning large genomic regions with often hundreds of genes. Fine-mapping an eQTL to identify the causal gene, or even polymorphic site, is not a trivial task.

If the same eQTL is found for several genes, a so-called 'hot-spot', this locus is likely to harbour a master regulator (Breitling *et al.*, 2008). Additionally, as the differences in gene expression underlie many phenotypic and physiological differences between genotypes, QTLs can be shared between genes and other quantitative traits (Fu *et al.*, 2009; van Zanten *et al.*, 2009; Burow *et al.*, 2010; Kerwin *et al.*, 2011, 2015; Joosen *et al.*, 2013b; He *et al.*, 2014; Jimenez-Gomez, 2014; Jensen *et al.*, 2015a,b). Moreover when gene expression is measured in multiple populations, plant stages, tissues or environments, the variation in these traits can be placed in a gene-by-environment context. This allows us to study why a trait displays variation in one environment and not in another, or why a trait shows different levels of variation within populations originating from crosses between different parental genotypes. The eQTL approach also laid the foundation for systems genetics, studying the interactions between traits and various layers of cellular organization (mRNA, protein, metabolite) by exploiting natural variation (Fu *et al.*, 2009; Civelek and Lusis, 2014).

To date, six original eQTL studies in RIL populations of *Arabidopsis thaliana* have been published (Keurentjes *et al.*, 2007; West *et al.*, 2007; Cubillos *et al.*, 2012; Snoek *et al.*, 2012; Joosen *et al.*, 2013b; Lowry *et al.*, 2013), together providing eight data sets of mapped eQTLs for a large number of Arabidopsis genes. However, these rich data sets with thousands of genetic interactions remain largely inaccessible to biologists who would like to analyse the interactions of their genes of interest to generate new hypotheses. Although several repositories exist that offer online access to QTL and eQTL data for various species, such as WormQTL (http://www.WormQTL.org) (Snoek *et al.*, 2013, 2014; van der Velde *et al.*, 2014), QTLstore (http://qtlstore.versailles.inra.fr/) (Cubillos *et al.*, 2012), Animal QTLdb (http://www.animalgenome.org/cgi-bin/QTLdb/index) (Hu *et al.*, 2016) and GeneNetwork (http://www.genenetwork.org) (Wu *et al.*, 2004), a versatile analysis platform with all Arabidopsis eQTL data was still lacking.

AraQTL

We address this problem with AraQTL (http://www.bioinformatics.nl/AraQTL/). AraQTL stores and combines all published Arabidopsis eQTL data and allows for investigations over different experiments. This greatly extends the possibilities for using eQTL data in on-going and future research. With a few clicks or a simple query, all genes with an eQTL at a specific marker or genomic region can be found, investigated and downloaded. Groups of genes with correlating eQTL patterns can be extracted from several different experiments and eQTL profiles from different populations can be compared. Moreover the outputs are interactively linked within AraQTL and to The Arabidopsis Information Resource (http://www.arabidopsis.org). Together, the ease of access to the data and possibilities for interactive exploration will facilitate system-level investigations, as well as follow up wet-lab experiments and importantly enable re-use of previously published data.

RESULTS

eQTL studies in AraQTL

AraQTL is a web-based workbench and database for investigation of eQTLs (Figure 1). It includes publicly available data from the six original eQTL studies in segregating populations of *A. thaliana* published to date (Table 1) (Keurentjes *et al.*, 2007; West *et al.*, 2007; Cubillos *et al.*, 2012; Snoek *et al.*, 2012; Joosen *et al.*, 2013a,b; Lowry *et al.*, 2013). Together, these studies provide eight sets of mapped eQTLs covering five different populations and are genotyped using over 600 specific genetic markers cumulatively. These populations are defined by their parental ecotypes: L*er* × Cvi, Bay × Sha, Cvi × Col, Bur × Col and Tsu

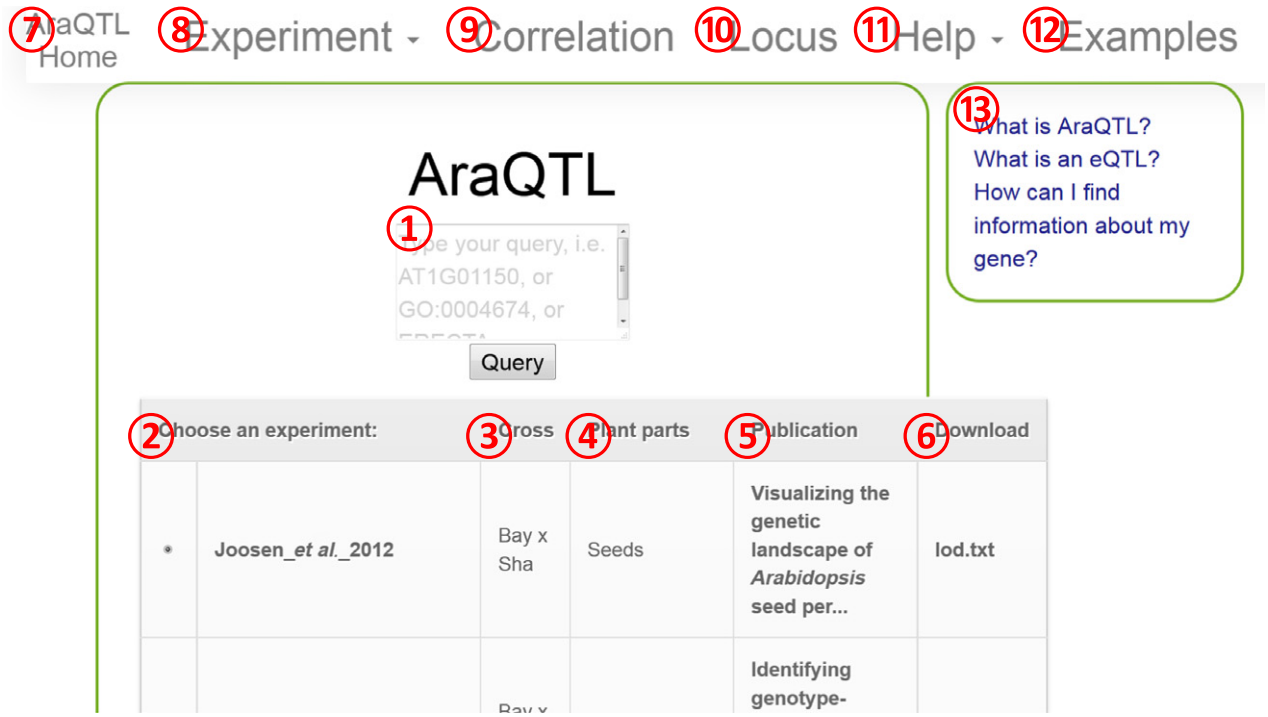

Figure 1. AraQTL homepage.
Different investigations can be started from this page. It contains several key features: ① The main search box, from which genes, markers, gene descriptions and Gene Ontology (GO) terms can be found. Searching for a single gene shows the expression quantitative trait locus (eQTL) profile for that gene for the selected experiment or all experiments when selected. Below this, the summary table lists datasets currently in AraQTL which consists of: ② data set selection, ③ recombinant inbred population used, ④ plant parts from which gene expression was measured, ⑤ title and link to the publication the dataset was originally generated for, and ⑥ a link to download the dataset with the eQTL profiles in simple txt format. At the top of the page a navigation bar can be found from which several selections can be made: ⑦ the link to the homepage, ⑧ a link to a genome-wide overview by *cis*/*trans* plot per experiment, ⑨ a link to eQTL profile correlation investigation, ⑩ a link to find all eQTLs at a specific locus, ⑪ a link to help features like, About, the manual and FAQ, and ⑫ a link to a page with clickable graphical examples of the different investigations. At ⑬ quick links to important information can be found.

× Kas. Gene expression was measured for over 20 000 different transcripts in more than 1200 samples taken from several different environments, growth stages and plant parts, such as seeds, seedlings, whole rosettes or specific leaves.

The first two genetical genomics studies in Arabidopsis were published in 2007. Both Keurentjes *et al.* (2007) and West *et al.* (2007) describe ample variation in gene expression between RILs of the L*er* × Cvi and Bay × Sha populations, respectively. For many genes (about 20–70%) the variation in gene expression could be explained (partially) by eQTLs (Keurentjes *et al.*, 2007; West *et al.*, 2007). In Keurentjes *et al.* (2007) the detected eQTLs were used to construct a gene regulatory network underlying flowering time, suggesting interesting new interactions. Both studies were used in many follow-up investigations to learn more about the regulation of specific genes (Terpstra *et al.*, 2010; Kerwin *et al.*, 2011; Joseph *et al.*, 2014; Li and Kliebenstein, 2014; Burow *et al.*, 2015; Leskow *et al.*, 2016) and gene expression networks underlying different traits (Kliebenstein *et al.*, 2006; Baginsky *et al.*, 2010; Kerwin *et al.*, 2011, 2015). This was also combined with variation of other molecular phenotypes such as protein abundance (Baginsky *et al.*, 2010) and metabolites (Keurentjes *et al.*, 2006).

In 2012 and 2013 four more eQTL studies in Arabidopsis were published (Cubillos *et al.*, 2012; Snoek *et al.*, 2012; Joosen *et al.*, 2013b; Lowry *et al.*, 2013). In Cubillos *et al.* (2012) the variation in gene expression and eQTLs of two populations sharing one parental strain were compared using rosette samples, showing an only moderately conserved eQTL landscape between the two populations. In three studies (Snoek *et al.*, 2012; Joosen *et al.*, 2013a,b; and Lowry *et al.*, 2013) eQTLs from different environments were compared. They showed that a large number of the eQTLs, especially the *trans*-eQTLs, are dependent on the environment and that gene-by-environment interaction is widespread for gene expression.

Even though many different hotspots (loci) for eQTLs are detected among and between populations, tissues and environments have only been detected for a few of the causal genes or regulatory sequence elements. One of the best studied is the *ERECTA* locus which has many eQTLs mapping to it in the L*er* × Cvi population (Keurentjes *et al.*,

Table 1 Studies available in AraQTL

Study	Population	Micro-array type	eQTL and thresholds[a]	Stage	Environment
Keurentjes *et al.* (2007)	L*er* × Cvi, 160 RILs	Qiagen AROSv1	G 3.83 (FDR 0.05)	7-day-old seedlings	Petri dish, growing conditions with 16 h light (30 W m–2) at 20°C, 8 h dark at 15°C and 75% relative humidity
West *et al.* (2007)	Bay × Sha, 211 RILs	Affymetrix ATH1 GeneChip	G 2.88 (FDR 0.05)	6-week-old rosette leaves	Growth chambers under short-day conditions (8 h light at 100–120 μEi, 20°C day/20°C night)
Cubillos *et al.* (2012)	Cvi × Col, 158 RILs Bur × Col, 156 RILs Bay × Sha, 157 RILs	CATMAv5 (Cvi × Col and Bur × Col), CATMAv2 (Bay × Sha)	G 3 (FDR 0.05)	Rosette developmental stage: Boyes 1.06 (Cvi × Col and Bur × Col) Developing seeds,10 days after pollination (Bay × Sha)	Greenhouse, with temperatures 20°C during the day and 15°C at night; long-day conditions (16-h days)
Snoek *et al.* (2012)	L*er* × Cvi, 120 RILs	CATMAv2	G, (G×E) 3.62 (FDR 0.05)	Rosette stage, leaves 24 days after germination	Growth chamber, under short days (9 h light), a light intensity of 200 μmol m^{-2} sec^{-1} at 20°C. Shade treatment 20 μmol m^{-2} sec^{-1} for 3 h
Joosen *et al.* (2013b,a)	Bay × Sha, 165 RILs	Affymetrix Arabidopsis SNPtile array	G, G×E	Seeds; bulk harvested and after-ripened. Germination started by 6 h imbibition.	In a climate chamber (20°C day, 18°C night) with 16 h of light (35 W m^{-2}) at a relative humidity of 70%
Lowry *et al.* (2013)	Tsu × Kas, 2 × 108 RILs	Affymetrix Arabidopsis SNPtile array	G, G×E 2.1 (FDR 0.1)	Rosette stage leaves (4 weeks) for, Wet and dry treatments.	Growth chamber, day length: 12 h. Light period: 23°C and 40% humidity. Dark period: 20°C and 50% humidity. Light intensity: 330 μmol m^{-2} sec^{-1}

eQTL, expression quantitative trait locus; RIL recombinant inbred line; FDR, false discovery rate
[a]eQTLs: G, genetic eQTLs; G×E, environment-specific eQTLs.

2007; van Zanten *et al.*, 2009; Terpstra *et al.*, 2010; Snoek *et al.*, 2012). Using several selection steps and enrichment tests based on eQTLs, transcription factor-binding sites, Gene Ontology (GO) terms and co-expression, a map kinase signalling cascade and a group of transcription factors from the *WRKY* family were found to be operating downstream of *ERECTA* (Terpstra *et al.*, 2010). Another elegant example of exploiting eQTL datasets to identify positional and functional candidate genes for phenotypic QTLs was published by Jimenez-Gomez *et al.* (2010): combining co-expression, eQTLs (West *et al.*, 2007) and functional annotation to efficiently reveal *ELF3* as the causal gene for a shade-avoidance response QTL in the Bay × Sha RIL population. These examples show that combining and using previously published eQTL data can lead to new insights and targeted experimental setups.

Getting started using the homepage and navigation bar

The homepage of AraQTL offers an overview of the studies and datasets currently hosted and the navigation tools for investigation and comparison (Figure 1). The online manual (http://www.bioinformatics.nl/AraQTL/?mode=manual) provides a detailed and user-friendly description of how to navigate through the database. The search box can be used to find the eQTL profiles for one or more genes in one or all experiments. Moreover, GO terms can be used to find the eQTL profiles of all genes annotated with that GO term. Any query term not directly linked to a gene ID

or GO term will report the genes and GO terms with that query term in their description. Investigation and exploration of the data is divided over different functions and pages, all interactively linked (Figure 2).

Experiments to be visualised can be pre-selected from the experiment overview table. This table also offers some basic information such as population used, plant tissues and original publication. The publication titles are linked to their PubMed pages for easy access and the data hosted in AraQTL can be downloaded in flat text format, for instance for further analysis with programming languages such as R or Python.

The navigation bar provides the main form of interactivity (Figure 2) and can be used for selection of graphical overviews, investigations and information. From left to right: the 'AraQTL' button enables quick return to the home page and search box; the 'Experiment' option can be used to go to a genome-wide overview for each experiment; the 'Correlation' option can be used to find genes with correlating eQTL profiles per experiment; the 'Locus' option can be used to find all genes with an eQTL at a specific marker or genomic coordinate; and 'Help' offers information about AraQTL, a link to the Manual and 'Frequently asked questions'. Through 'Examples' the main investigation options can be selected via a visual and clickable overview, to quickly explore the different functions of AraQTL.

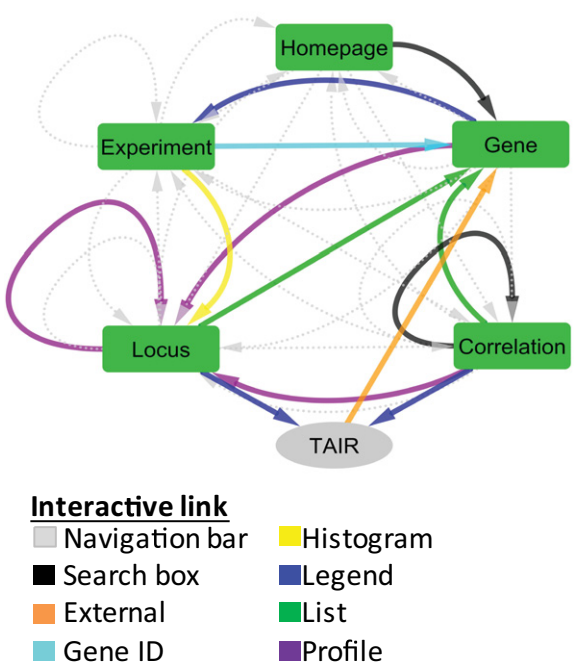

Figure 2. Contact map of AraQTL.
The legend shows the interactive links and which parts can be selected to go further with an investigation within AraQTL (green nodes) or through The Arabidopsis Information Resource (TAIR; grey node). Dotted grey arrows show the links via the top navigation bar. Black arrows show connections via the search box. Dark blue lines represent the connections through the figure legend. Light blue arrows represent the links via the plot title. Yellow arrows show the connections via the expression quantitative trait locus (eQTL) distribution histogram. Purple arrows indicate the connections through the eQTL profile and green arrows show the connections by the gene list below the eQTL profile figures.

Benefits of using AraQTL

AraQTL was developed to unlock the potential of eQTL datasets for further research. The major feature which sets AraQTL apart from other databases is that it enables groups of genes to be selected based on a shared genetic effect on their expression. To enable direct comparisons of eQTLs between experiments the genetic maps of the different populations have been integrated based on physical marker positions. AraQTL uniquely allows researchers to identify whether a group of genes (defined by a specific GO term) have a shared eQTL hotspot. This is followed by integrative tools for finding which other genes have co-locating eQTLs and exporting those as a list to an external analysis platform. Altogether, groups of genes with a shared genetic architecture can be very easily investigated within and beyond AraQTL, making it a very versatile tool for plant scientists.

Example 1: starting an investigation with the GO term 'Detection of visible light'. The search box can be used to find one or more genes by ID or by name, but also to find GO terms. When a specific GO term ID is entered, the eQTL profiles for all genes annotated with that GO term will be shown. When (part of) a GO term description is entered, all genes and GO terms with that description will be displayed in a list from which they can be selected for investigation (one at a time). GO terms can be quickly found by clicking 'GO terms' at the top of the list. The experiment selection box in the navigation bar can be used to switch between different experiments. Multiple browser windows can be used to compare the eQTL profiles from different experiments.

For example, to investigate natural variation in gene expression associated with 'light', the search box can be used (Figure 3a). Clicking the 'Query' button results in a list containing 45 genes and 90 GO terms (Figure 3b). A GO term can be selected for further investigation by clicking its GO ID. When we select the GO term GO:0009584 'Detection of visible light' (Figure 3c) the eQTL profiles of the five genes annotated with this GO term, phytochromes *PHYA* to *PHYE* are shown. The eQTLs from the different experiments can be selected from the navigation bar. When the experiment Snoek_Terpstra_etal_2012 is selected we can observe that *PHYB* and *PHYE* both have a potential *cis*-eQTL (an eQTL mapping close to its own genomic location; Figure 3d). *PHYE* has a second eQTL on chromosome 5 which indicates *trans*-regulation.

Other experiments show different eQTL profiles for the phytochromes and can be compared by opening multiple

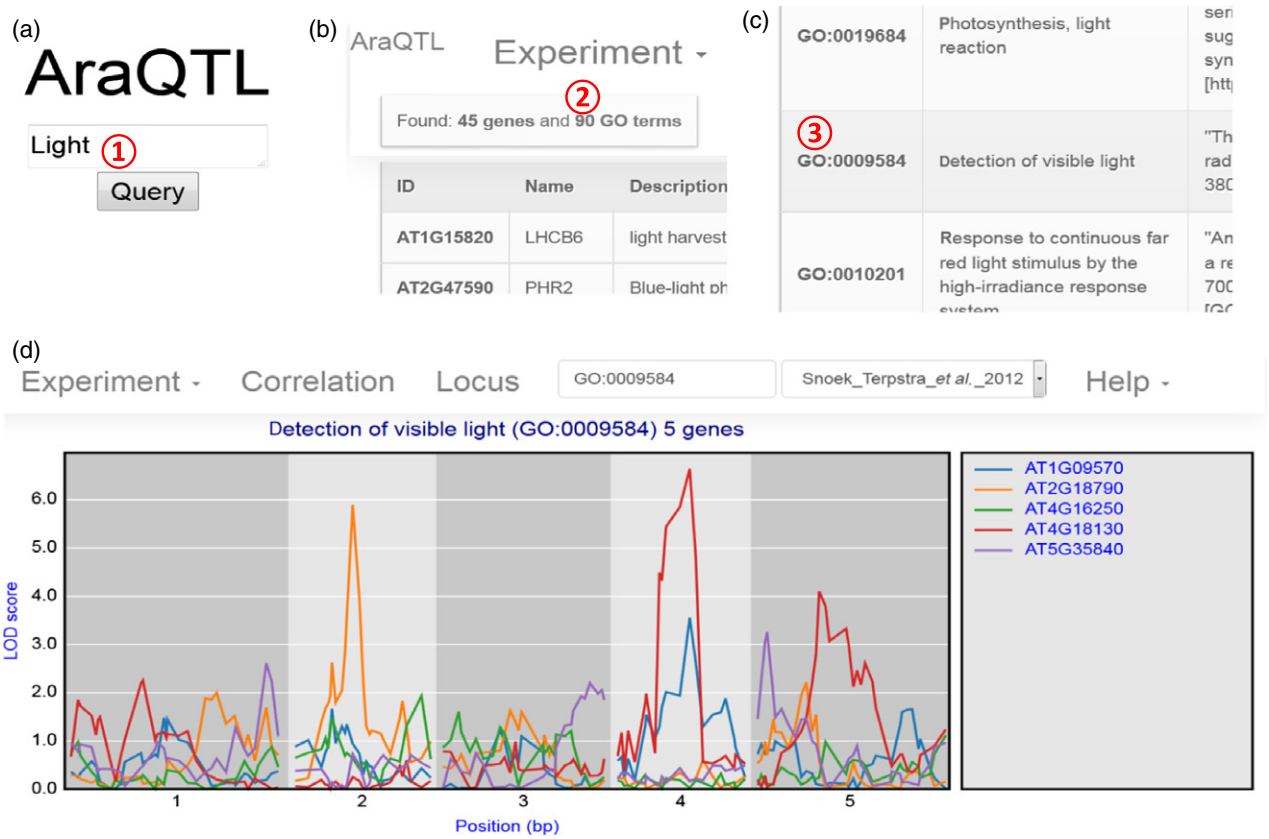

Figure 3. Gene Ontology (GO) term-specific expression quantitative trait locus (eQTL) profiles.
(a) Search box entry to find GO terms related to 'light'.
(b) Top of the results page. By clicking 'GO terms' the list jumps to GO terms.
(c) Description and selection of GO terms, by clicking the GO ID number the eQTL profiles are shown.
(d) eQTL profiles of the selected GO term. The different experiments can be selected from the dropdown menu from the navigation bar.

browser windows. This can be done by clicking the gene IDs in the list below the figure, which will display the gene-specific eQTL profiles found in all experiments in AraQTL for a single gene. Comparing the gene-specific eQTL profiles of all five phytochromes we find that four, *PHYB* (*At2g18790*), *PHYC* (*At5g35840*), *PHYD* (*At4g16250*) and *PHYE* (*At4g18130*), have a local, potential *cis*-eQTL in at least one of the experiments (Figure S1 in the Supporting Information), while *PHYA* (*At1g09570*) only shows three minor *trans*-eQTLs. Most phytochrome eQTL profiles differ between the experiments. This suggests gene-by-environment interaction or gene-by-genotype interaction (epistasis) for the expression of the phytochromes and cryptic genetic variation (CGV), i.e. variation that, under atypical environmental conditions or specific genetic backgrounds, generates heritable phenotypic variation and induces evolution (Paaby and Rockman, 2014). Such gene-by-environment or epistatic interactions, where allelic effects are dependent on the environment, may be important for a range of phenotypic outcomes. AraQTL allows for the integration of environmental conditions and crosses to investigate complex interactions and CGV at the level of the transcriptome and subsequent eQTL mappings. This provides new insights into the genetic architecture of environment-dependent traits.

Gene expression variation at the eQTL loci of the phytochromes can be further explored by clicking one of the phytochrome eQTL peaks. This shows those genes with a co-locating eQTL. When we click the potential *cis*-eQTL peak of *PHYB* at marker BF.221L found in the Snoek_Terpstra_etal_2012 experiment we find 599 genes with a co-locating eQTL [significance threshold logarithm of the odds (LOD) > 3.1]. To exclude the potential *cis*-eQTLs we can select an exclusion window in the navigation bar. Excluding all genes at a distance of less than 5 Mb from the eQTL leaves 376 genes. These genes have co-locating *trans*-eQTLs and are potentially co-regulated. Genes with a potential *cis*-eQTL at this locus, like *PHYB*, are probable candidates. Further study is needed to find which genes are regulated by *PHYB*; most likely there are also other polymorphic regulators at this locus.

Example 2: genome-wide overview and exploration of an eQTL hotspot. To get a genome-wide overview of the

distribution of eQTLs, a *cis*/*trans* plot (Figure 4) is most often used. The position of each eQTL peak is plotted on the *x*-axis against the positions of the genes for which this eQTL peak was found. In this plot, several patterns can be detected. The diagonal line shows the likely *cis*-eQTLs, which are those eQTLs mapping close to the position of the gene for which they have been found. Vertical bands ('*trans*-bands') can also be observed; these each contain many eQTLs, which map to the same (small) genomic region in which they themselves are not physically located. These *trans*-band loci affect the expression of many genes and are likely to contain one or more key regulators which act on transcription either directly or indirectly through some other phenotypes.

A default eQTL peak significance threshold per experiment is provided for quick visualisation, and can be manually adjusted for further exploration. The *cis*/*trans* plot is interactive, allowing selection of different experiments via a drop-down menu. Moreover the individual significant peaks (points in the *cis*/*trans* plot) can be clicked to show the complete eQTL profile of the selected gene/peak. Finally, the total peak count per marker is indicated in the bar plot below the *cis*/*trans* plot. Individual bars/markers can be clicked to obtain a figure and list with genes that share a peak at the selected marker.

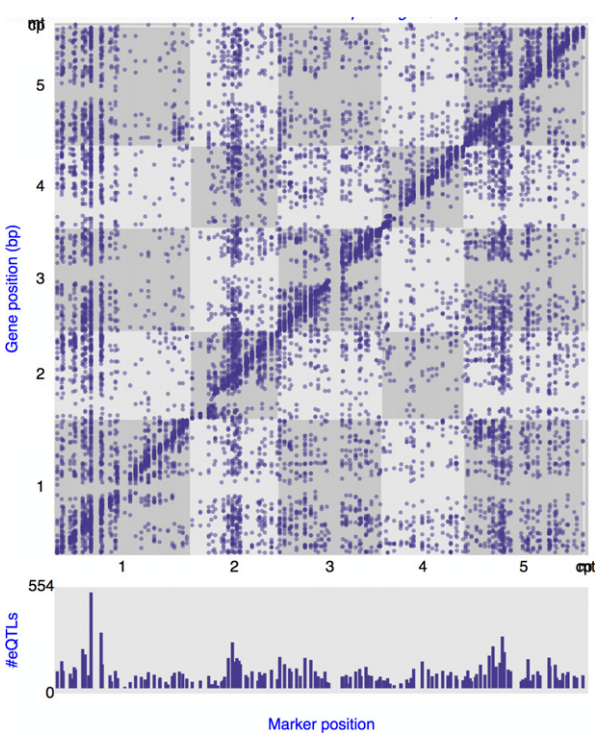

Figure 4. *Cis*/*trans* plot.
This page gives a genome-wide overview per experiment of all the detected expression quantitative trait loci in the Snoek_Terpstra_etal_2012 experiment. The major *trans*-band is located on chromosome 1.

In the genome-wide overview of Snoek *et al.* (2012) we can see a *trans*-band on chromosome 1 at marker CH.160L (Figure 4). Hovering over the bar in the histogram shows that 554 genes have an eQTL peak at this locus. Clicking the bar shows the complete eQTL profiles of all genes mapping to this locus in the 'Locus' window. Here we can select those genes being regulated from this locus *in trans*, by an exclusion window of 5 Mb in the navigation bar. The 645 genes with a significant *trans*-eQTL (LOD > 3.1) can be explored one by one or compared with genes with a potential *cis*-eQTL at this locus. For further investigation the list of 645 genes can be copied from the 'Download gene list' option for use in other online tools such as AgriGO (http://bioinfo.cau.edu.cn/agriGO/) (Du *et al.*, 2010) or AraNetv2 (http://www.inetbio.org/aranet/) (Lee *et al.*, 2010).

AgriGO can be used, like other web-based GO enrichment tools, to find the functional consequences of the genes being affected by the eQTL. The AgriGO Singular Enrichment Analysis (SEA) tool (Table S1) shows that the 645 genes selected are enriched for GO terms like 'Circadian rhythm' and 'Nitrogen compound metabolic process'; moreover, many genes show 'Transcription regulator activity', possibly explaining why so many genes map to this locus. A candidate gene for affecting the expression of all these genes and subsequent processes is *GIGANTEA* (*GI*). The genomic location of *GI* is very close to marker CH.160L, and is involved as a master regulator in the circadian rhythm process. Moreover it is involved in flowering time and it has previously been suggested that it is involved in expression level variation mapping to this locus (Keurentjes *et al.*, 2007).

AraNet can be used to place a set of genes in a probabilistic functional network based on published data and co-expression studies (Lee *et al.*, 2010). A useful feature here is the test for enrichment in linkages between the selected genes compared with a random set of genes. When significant, this is strong evidence that those genes with co-locating eQTLs are functioning together in a pathway or process. When the selected set of 645 genes are further investigated using the function 'Find new members of a pathway', we find these genes to be significantly more interconnected than a random set of genes ($P < 1.8 \times 10^{-6}$; Figure S2). They are therefore very likely to operate in one or a few closely linked pathways. Using the connections in AraNet, the relation between the genes with co-locating *trans*-eQTLs can be studied (Figure S3). Furthermore, AraNet provides a list of genes that are strongly associated with the set of genes sharing an eQTL locus. These strongly associated genes are candidate regulators and can be further selected by their genomic positions as the location of the eQTLs is known.

Together these are the first steps to a system genetics approach; gene expression network generation and refinement, followed by annotation enrichment providing the

context for the next targeted experiment to show gene regulation and function. The first steps have been made interactive and will be further automated in the near future. These examples demonstrate how eQTL data can already be explored relatively easy to discover new groups of interacting genes by using AraQTL in a web based workflow.

DISCUSSION

AraQTL is designed to improve the accessibility of Arabidopsis eQTL data, serving a wide range of scientists using Arabidopsis as a model. Storing eQTL data in one platform in a single, unified format provides an easy way of visually inspecting eQTL patterns. These patterns are directly useful for analysis because they indicate whether a gene varies in expression and which loci are likely to be affecting its expression levels. Very often the underlying papers do not include the eQTL profiles as supporting information [but see Cubillos *et al.* (2012) in QTLstore; http://qtlstore.versailles.inra.fr]. Furthermore, the original publications seldom include all the required data, tools and methods required to re-map the eQTLs, such as normalised expression data and genotypes or mapping models and scripts. Having the results of these datasets, the eQTL profiles, together in AraQTL facilitates integration and allows for comparisons which would otherwise be cumbersome and laborious.

Even though all eQTL sets can now be explored in a combined way, some limitations manifest themselves when comparing them. This mainly has to do with the different technologies used in these experiments to measure gene expression. Early microarray technology used probes spotted on a glass slide, which may lead to false positives and false negatives (Alberts *et al.*, 2007). The absence of a probe prevents the interrogation for that gene and closely related genes can cross-hybridize due to probe similarity. Moreover, eQTLs could be mapped where transcript abundance variation is misrepresented by differences in hybridisation due to sequence polymorphisms. This effect is stronger with shorter probes, as for example used on Affymetrix arrays, but is certainly not absent from arrays using longer probes. Consequently, the end-user should be cautious when comparing results from different populations. Some genes may be absent from one specific population, not because of lack of variation, but due to lack of interrogation or because their variation is diluted due to a non-specific probe (false negative).

On another note, the mapping power in different populations may vary, since the number and distribution of recombination sites between the RILs as well as the number of markers used to genotype them are different between the sets, even though the type and number of RILs used in the current datasets is rather uniform. AraQTL hosts the eQTL profiles as they were originally mapped, which means that differences in specific models and mapping procedures could give rise to some of the differences between the different sets of eQTLs. Non-linear relationships between genetic and physical maps in specific crosses may also influence the comparison of QTL localization.

Future directions

The aim of AraQTL is to make the data needed for re-mapping easily accessible, so eQTLs from different experiments can be detected by the same method and therefore better compared. This allows alternative models to be explored or expression levels of different experiments to be included in one eQTL mapping model. Genetic maps can also be improved by including gene expression markers (West *et al.*, 2006) or through sequencing. This will lead to eQTLs with a higher resolution and better regulatory prediction. In future updates of AraQTL these functions will be implemented, yet easy access to the data already enables an efficient start so that anyone can for further exploit eQTL and other system genetics data.

Combining established high-throughput measurement techniques like next generation sequencing, proteomics, metabolomics and phenomics offers great potential for plant sciences and quantitative genetics (Figure 5). These

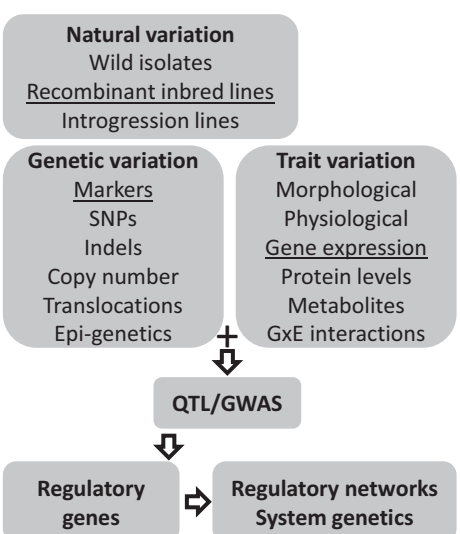

Figure 5. From natural variation to regulatory networks.
Natural variation found in mapping populations like wild isolates, recombinant inbred lines and introgression lines offers a versatile resource to link genetic variation to trait variation. Both genetic and trait variation come in several different forms. For example, single nucleotide polymorphisms (SNPs), indels, copy number, translocations and even epi-genetics can be potential causal factors of trait variation. Trait variation can be measured on many different levels – morphological, physiological, gene expression, protein and metabolites. This variation often interacts with changing environmental conditions. Through quantitative trait locus mapping or genome-wide association studies, regulatory loci and genes can be detected. By using a genome-wide and systematic approach regulatory networks can be constructed, followed by a system genetics approach in which all levels of variation are combined to learn the true nature of the genetic architecture of traits. The underlined factors are those which have been used to generate the expression quantitative trait loci currently stored in AraQTL. The remaining factors are expansion possibilities to host and connect through AraQTL.

methods generate a wealth of data, which makes the storage, access and especially the generation of useful and meaningful connections within and between the different types of data increasingly important. Moreover, results from different types of mapping populations can be included. In this way the advantages of introgression line populations, RIL populations, multi-parental mapping panels and sets of wild isolates can be combined. With this in mind, the next steps for AraQTL will be linking eQTL data to polymorphisms from massive sequencing projects of many different ecotypes (Gan *et al.*, 2011; Genomes Consortium. Electronic address and Genomes, 2016) and including eQTLs and single nucleotide polymorphisms (SNPs) obtained from RNA-seq experiments. When stored in, and visualised from, the same platform, the SNPs and phenotypes enable the integration of QTL mapping and GWAS investigation, further increasing the detection power of both methods. For example, eQTL datasets have been successfully combined with results from transcriptomic GWAS (Kawakatsu *et al.*, 2016) and allele-specific expression (Cubillos *et al.*, 2014) RNA-seq experiments.

New tools for investigation and visualisation will be developed in a modular fashion for easy integration and deployment within AraQTL. The annotation side can be expanded beyond GO terms, for example with pathway knowledge – as available through the Kyoto Encyclopaedia of Genes and Genomes (KEGG) (http://www.genome.jp/kegg/) or with gene association networks like AraNet (Lee *et al.*, 2010). To further investigate the relation between genotype and variation in phenotype, AraQTL will be expanded with published and new classical/phenotypic QTLs. This enables searching for the possible molecular components underlying a QTL for a specific phenotype and finding the possible causal genes. Combining highly detailed molecular data, such as transcription factor-binding and histone-binding sites or protein–protein interactions, will allow for even more powerful analyses.

AraQTL has been designed from the start to easily store and share upcoming RNA-seq data and eQTLs from these data and QTLs from metabolomics and proteomics and to visualise and analyse these together. Comparing sets of genes through functional enrichment will enable an even better, more targeted, approach to candidate gene selection and network generation to link gene expression, genetic variation and function. In the near future, tools will be developed to investigate genetic variation in a more systematic genome- and population-wide manner, enabling more complete and higher-resolution system genetics.

We believe the eQTL data in AraQTL will greatly benefit the Arabidopsis research community, providing a rich source of genetic interactions to plant biologists. AraQTL will serve as a solid platform for in-depth analysis of these interactions to help chart the Arabidopsis gene regulatory network.

EXPERIMENTAL PROCEDURES

eQTL profiles mapped in the original papers were used for visualization and investigation, except for those found in Lowry *et al.* (2013) which were re-mapped from the original genotypes and gene expression data by a single marker model using a linear model in R. Original determined genome-wide threshold levels may be applied to call significant eQTLs (Table 1). The thresholds for comparing eQTLs will vary depending on the number of genes, eQTLs and populations involved. The physical positions of the markers were used to compare and integrate the genetic maps of the different populations and can be obtained from the AraQTL website.

All pairwise correlations between eQTL patterns were calculated using the Pearson correlation coefficients between the eQTL patterns of genes within an experiment using the R function 'cor', on the LOD scores. Searching for genes with an eQTL at a specific locus is implemented by selecting genes that have a LOD score above the given threshold at the marker closest to the specified locus. To select *trans*-eQTLs genes with a *cis*-eQTL can be excluded based on their physical distance to this marker.

GO terms were downloaded from The Arabidopsis Information Resource (TAIR10) and gene descriptions from Ensembl BioMart (Kinsella *et al.*, 2011).

AraQTL was developed using the Python Django web framework. The backend runs on an Ubuntu 16.04.1 LTS Linux server, using the Apache web server version 2.4.18 and a MySQL 5.7.13 database. The web frontend is implemented via Django templates in HTML and Javascript, using the D3 library and Jquery. The *cis*/*trans* plot and QTL profile plots are built upon work by Karl Broman (Broman, 2015).

ACKNOWLEDGEMENTS

We thank Karl Broman for sharing his D3 QTL plotting scripts. BLS was funded by the Netherlands Organisation for Scientific Research (project no. 823.01.001). The authors declare no conflict of interest.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Expression quantitative trait locus profiles of the phytochromes.

Figure S2. Results from Aranetv2.

Figure S3. Results from Aranetv2 in a network visualisation.

Table S1. AgriGO results.

REFERENCES

Alberts, R., Terpstra, P., Li, Y., Breitling, R., Nap, J.P. and Jansen, R.C. (2007) Sequence polymorphisms cause many false cis eQTLs. *PLoS ONE*, **2**, e622.

Alonso-Blanco, C. and Koornneef, M. (2000) Naturally occurring variation in Arabidopsis: an underexploited resource for plant genetics. *Trends Plant Sci.* **5**, 22–29.

Alonso-Blanco, C., Aarts, M.G., Bentsink, L., Keurentjes, J.J., Reymond, M., Vreugdenhil, D. and Koornneef, M. (2009) What has natural variation taught us about plant development, physiology, and adaptation? *Plant Cell*, **21**, 1877–1896.

Baginsky, S., Hennig, L., Zimmermann, P. and Gruissem, W. (2010) Gene expression analysis, proteomics, and network discovery. *Plant Physiol.* **152**, 402–410.

Breitling, R., Li, Y., Tesson, B.M. *et al.* (2008) Genetical genomics: spotlight on QTL hotspots. *PLoS Genet.* **4**, e1000232.

Brennan, A.C., Mendez-Vigo, B., Haddioui, A., Martinez-Zapater, J.M., Pico, F.X. and Alonso-Blanco, C. (2014) The genetic structure of Arabidopsis thaliana in the south-western Mediterranean range reveals a shared history between North Africa and southern Europe. *BMC Plant Biol.* **14**, 17.

Broman, K.W. (2015) R/qtlcharts: interactive graphics for quantitative trait locus mapping. *Genetics*, **199**, 359–361.

Burow, M., Halkier, B.A. and Kliebenstein, D.J. (2010) Regulatory networks of glucosinolates shape Arabidopsis thaliana fitness. *Curr. Opin. Plant Biol.* **13**, 348–353.

Burow, M., Atwell, S., Francisco, M., Kerwin, R.E., Halkier, B.A. and Kliebenstein, D.J. (2015) The glucosinolate biosynthetic gene AOP2 mediates feed-back regulation of jasmonic acid signaling in Arabidopsis. *Mol Plant*, **8**, 1201–1212.

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

Chan, E.K., Rowe, H.C., Corwin, J.A., Joseph, B. and Kliebenstein, D.J. (2011) Combining genome-wide association mapping and transcriptional networks to identify novel genes controlling glucosinolates in Arabidopsis thaliana. *PLoS Biol.* **9**, e1001125.

Civelek, M. and Lusis, A.J. (2014) Systems genetics approaches to understand complex traits. *Nat. Rev. Genet.* **15**, 34–48.

Cubillos, F.A., Yansouni, J., Khalili, H. *et al.* (2012) Expression variation in connected recombinant populations of Arabidopsis thaliana highlights distinct transcriptome architectures. *BMC Genom.* **13**, 117.

Cubillos, F.A., Stegle, O., Grondin, C., Canut, M., Tisne, S., Gy, I. and Loudet, O. (2014) Extensive cis-regulatory variation robust to environmental perturbation in Arabidopsis. *Plant Cell*, **26**, 4298–4310.

Du, Z., Zhou, X., Ling, Y., Zhang, Z. and Su, Z. (2010) agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res.* **38**, W64–W70.

Fu, J., Keurentjes, J.J., Bouwmeester, H. *et al.* (2009) System-wide molecular evidence for phenotypic buffering in Arabidopsis. *Nat. Genet.* **41**, 166–167.

Gan, X., Stegle, O., Behr, J. *et al.* (2011) Multiple reference genomes and transcriptomes for Arabidopsis thaliana. *Nature*, **477**, 419–423.

Genomes Consortium. Electronic address, m.n.g.o.a.a. and Genomes Consortium (2016) 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. *Cell*, **166**, 481–491.

Gomaa, N.H., Montesinos-Navarro, A., Alonso-Blanco, C. and Pico, F.X. (2011) Temporal variation in genetic diversity and effective population size of Mediterranean and subalpine Arabidopsis thaliana populations. *Mol. Ecol.* **20**, 3540–3554.

He, H., de Souza Vidigal, D., Snoek, L.B., Schnabel, S., Nijveen, H., Hilhorst, H. and Bentsink, L. (2014) Interaction between parental environment and genotype affects plant and seed performance in Arabidopsis. *J. Exp. Bot.* **65**, 6603–6615.

Hu, Z.L., Park, C.A. and Reecy, J.M. (2016) Developmental progress and current status of the Animal QTLdb. *Nucleic Acids Res.* **44**, D827–D833.

Jansen, R.C. (2003) Studying complex biological systems using multifactorial perturbation. *Nat. Rev. Genet.* **4**, 145–151.

Jansen, R.C. and Nap, J.P. (2001) Genetical genomics: the added value from segregation. *Trends Genet.* **17**, 388–391.

Jensen, L.M., Jepsen, H.S., Halkier, B.A., Kliebenstein, D.J. and Burow, M. (2015a) Natural variation in cross-talk between glucosinolates and onset of flowering in Arabidopsis. *Front Plant Sci.* **6**, 697.

Jensen, L.M., Kliebenstein, D.J. and Burow, M. (2015b) Investigation of the multifunctional gene AOP3 expands the regulatory network fine-tuning glucosinolate production in Arabidopsis. *Front Plant Sci.* **6**, 762.

Jimenez-Gomez, J.M. (2014) Network types and their application in natural variation studies in plants. *Curr. Opin. Plant Biol.* **18**, 80–86.

Jimenez-Gomez, J.M., Wallace, A.D. and Maloof, J.N. (2010) Network analysis identifies ELF3 as a QTL for the shade avoidance response in Arabidopsis. *PLoS Genet.* **6**, e1001100.

Joosen, R.V., Ligterink, W., Hilhorst, H.W. and Keurentjes, J.J. (2009) Advances in genetical genomics of plants. *Curr. Genomics*, **10**, 540–549.

Joosen, R., Willems, L., Provart, N., Ligterink, W. and Hilhorst, H. (2013a) Next generation eQTL mapping; a sneak preview. *Imaging genetics of seed performance*, **111**, 111–126.

Joosen, R.V., Arends, D., Li, Y., Willems, L.A., Keurentjes, J.J., Ligterink, W., Jansen, R.C. and Hilhorst, H.W. (2013b) Identifying genotype-by-environment interactions in the metabolism of germinating arabidopsis seeds using generalized genetical genomics. *Plant Physiol.* **162**, 553–566.

Joseph, B., Atwell, S., Corwin, J.A., Li, B. and Kliebenstein, D.J. (2014) Meta-analysis of metabolome QTLs in Arabidopsis: trying to estimate the network size controlling genetic variation of the metabolome. *Front Plant Sci.* **5**, 461.

Kawakatsu, T., Huang, S.S., Jupe, F. *et al.* (2016) Epigenomic diversity in a global collection of Arabidopsis thaliana accessions. *Cell*, **166**, 492–505.

Kerwin, R.E., Jimenez-Gomez, J.M., Fulop, D., Harmer, S.L., Maloof, J.N. and Kliebenstein, D.J. (2011) Network quantitative trait loci mapping of circadian clock outputs identifies metabolic pathway-to-clock linkages in Arabidopsis. *Plant Cell*, **23**, 471–485.

Kerwin, R., Feusier, J., Corwin, J. *et al.* (2015) Natural genetic variation in Arabidopsis thaliana defense metabolism genes modulates field fitness. *eLife*, **4**, doi: 10.7554/eLife.05604.

Keurentjes, J.J. (2009) Genetical metabolomics: closing in on phenotypes. *Curr. Opin. Plant Biol.* **12**, 223–230.

Keurentjes, J.J., Fu, J., de Vos, C.H., Lommen, A., Hall, R.D., Bino, R.J., van der Plas, L.H., Jansen, R.C., Vreugdenhil, D. and Koornneef, M. (2006) The genetics of plant metabolism. *Nat. Genet.* **38**, 842–849.

Keurentjes, J.J., Fu, J., Terpstra, I.R., Garcia, J.M., van den Ackerveken, G., Snoek, L.B., Peeters, A.J., Vreugdenhil, D., Koornneef, M. and Jansen, R.C. (2007) Regulatory network construction in Arabidopsis by using genome-wide gene expression quantitative trait loci. *Proc. Natl Acad. Sci. USA*, **104**, 1708–1713.

Keurentjes, J.J., Koornneef, M. and Vreugdenhil, D. (2008a) Quantitative genetics in the age of omics. *Curr. Opin. Plant Biol.* **11**, 123–128.

Keurentjes, J.J., Sulpice, R., Gibon, Y., Steinhauser, M.C., Fu, J., Koornneef, M., Stitt, M. and Vreugdenhil, D. (2008b) Integrative analyses of genetic variation in enzyme activities of primary carbohydrate metabolism reveal distinct modes of regulation in Arabidopsis thaliana. *Genome Biol.* **9**, R129.

Keurentjes, J.J., Angenent, G.C., Dicke, M., Dos Santos, V.A., Molenaar, J., van der Putten, W.H., de Ruiter, P.C., Struik, P.C. and Thomma, B.P. (2011) Redefining plant systems biology: from cell to ecosystem. *Trends Plant Sci.* **16**, 183–190.

Kinsella, R.J., Kahari, A., Haider, S. *et al.* (2011) Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database: the journal of biological databases and curation*, **2011**, bar030. doi: 10.1093/database/bar030

Kliebenstein, D.J., West, M.A., van Leeuwen, H., Loudet, O., Doerge, R.W. and St Clair, D.A. (2006) Identification of QTLs controlling gene expression networks defined a priori. *BMC Bioinformatics*, **7**, 308.

Koornneef, M. and Meinke, D. (2010) The development of Arabidopsis as a model plant. *The Plant journal: for cell and molecular biology*, **61**, 909–921.

Lee, I., Ambaru, B., Thakkar, P., Marcotte, E.M. and Rhee, S.Y. (2010) Rational association of genes with traits using a genome-scale gene network for Arabidopsis thaliana. *Nat. Biotechnol.* **28**, 149–156.

Leskow, C.C., Kamenetzky, L., Dominguez, P.G. *et al.* (2016) Allelic differences in a vacuolar invertase affect Arabidopsis growth at early plant development. *J. Exp. Bot.* **67**, 4091–4103.

Li, B. and Kliebenstein, D.J. (2014) The AT-hook motif-encoding gene METABOLIC NETWORK MODULATOR 1 underlies natural variation in Arabidopsis primary metabolism. *Front Plant Sci.* **5**, 415.

Lowry, D.B., Logan, T.L., Santuari, L., Hardtke, C.S., Richards, J.H., DeRose-Wilson, L.J., McKay, J.K., Sen, S. and Juenger, T.E. (2013) Expression quantitative trait locus mapping across water availability environments reveals contrasting associations with genomic features in Arabidopsis. *Plant Cell*, **25**, 3266–3279.

Molenaar, J.A. and Keurentjes, J.J. (2014) Exploiting natural variation in Arabidopsis. *Methods Mol. Biol.* **1062**, 139–153.

Paaby, A.B. and Rockman, M.V. (2014) Cryptic genetic variation: evolution's hidden substrate. *Nat. Rev. Genet.* **15**, 247–258.

Rockman, M.V. and Kruglyak, L. (2006) Genetics of global gene expression. *Nat. Rev. Genet.* **7**, 862–872.

Snoek, L.B., Terpstra, I.R., Dekter, R., Van den Ackerveken, G. and Peeters, A.J. (2012) Genetical genomics reveals large scale genotype-by-environment interactions in Arabidopsis thaliana. *Frontiers in genetics*, **3**, 317.

Snoek, L.B., Van der Velde, K.J., Arends, D. *et al.* (2013) WormQTL–public archive and analysis web portal for natural variation data in Caenorhabditis spp. *Nucleic Acids Res.* **41**, D738–D743.

Snoek, L.B., van der Velde, K.J., Li, Y., Jansen, R.C., Swertz, M.A. and Kammenga, J.E. (2014) Worm variation made accessible: take your shopping cart to store, link, and investigate! *Worm*, **3**. doi: 10.4161/worm.28357.

Stitt, M., Sulpice, R. and Keurentjes, J. (2010) Metabolic networks: how to identify key components in the regulation of metabolism and growth. *Plant Physiol.* **152**, 428–444.

Taylor-Teeples, M., Lin, L., de Lucas, M. *et al.* (2015) An Arabidopsis gene regulatory network for secondary cell wall synthesis. *Nature*, **517**, 571–575.

Terpstra, I.R., Snoek, L.B., Keurentjes, J.J., Peeters, A.J. and van den Ackerveken, G. (2010) Regulatory network identification by genetical genomics: signaling downstream of the Arabidopsis receptor-like kinase ERECTA. *Plant Physiol.* **154**, 1067–1078.

Trontin, C., Tisne, S., Bach, L. and Loudet, O. (2011) What does Arabidopsis natural variation teach us (and does not teach us) about adaptation in plants? *Curr. Opin. Plant Biol.* **14**, 225–231.

van der Velde, K.J., de Haan, M., Zych, K., Arends, D., Snoek, L.B., Kammenga, J.E., Jansen, R.C., Swertz, M.A. and Li, Y. (2014) WormQTLHD–a web database for linking human disease to natural variation data in C. elegans. *Nucleic Acids Res.* **42**, D794–D801.

Vignes, M., Vandel, J., Allouche, D., Ramadan-Alban, N., Cierco-Ayrolles, C., Schiex, T., Mangin, B. and de Givry, S. (2011) Gene regulatory network reconstruction using Bayesian networks, the Dantzig Selector, the Lasso and their meta-analysis. *PLoS ONE*, **6**, e29165.

Weigel, D. (2012) Natural variation in Arabidopsis: from molecular genetics to ecological genomics. *Plant Physiol.* **158**, 2–22.

West, M.A., van Leeuwen, H., Kozik, A., Kliebenstein, D.J., Doerge, R.W., St Clair, D.A. and Michelmore, R.W. (2006) High-density haplotyping with microarray-based expression and single feature polymorphism markers in Arabidopsis. *Genome Res.* **16**, 787–795.

West, M.A., Kim, K., Kliebenstein, D.J., van Leeuwen, H., Michelmore, R.W., Doerge, R.W. and St Clair, D.A. (2007) Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in Arabidopsis. *Genetics*, **175**, 1441–1450.

Wijnen, C.L. and Keurentjes, J.J. (2014) Genetic resources for quantitative trait analysis: novelty and efficiency in design from an Arabidopsis perspective. *Curr. Opin. Plant Biol.* **18**, 103–109.

Wu, C.C., Huang, H.C., Juan, H.F. and Chen, S.T. (2004) GeneNetwork: an interactive tool for reconstruction of genetic networks using microarray data. *Bioinformatics*, **20**, 3691–3693.

van Zanten, M., Snoek, L.B., Proveniers, M.C. and Peeters, A.J. (2009) The many functions of ERECTA. *Trends Plant Sci.* **14**, 214–218.

Zhu, J., Zhang, B., Smith, E.N., Drees, B., Brem, R.B., Kruglyak, L., Bumgarner, R.E. and Schadt, E.E. (2008) Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. *Nat. Genet.* **40**, 854–861.