Next generation Genetical Genomics analysis

Wilco Ligterink, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands, phone: +31 317 482809, email: wilco.ligterink@wur.nl

Research Area: Breakthrough Technologies; Genome Analysis; Bioinformatics
Identifying genotype-by-environment interactions in the metabolism of germinating Arabidopsis seeds using Generalized Genetical Genomics

Ronny Viktor Louis Joosen*, Wageningen Seed Lab, Laboratory of Plant Physiology, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands

Danny Arends*, GBIC Groningen Bioinformatics Centre, University of Groningen, Nijenborgh 7, 9747 AG Groningen, The Netherlands

Yang Li*, GBIC Groningen Bioinformatics Centre, University of Groningen, Nijenborgh 7, 9747 AG Groningen, The Netherlands

Leo AJ Willems, Wageningen Seed Lab, Laboratory of Plant Physiology, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands

Joost JB Keurentjes, Laboratory of Genetics, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands; Laboratory of Plant Physiology, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands

Wilco Ligterink, Wageningen Seed Lab, Laboratory of Plant Physiology, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands

Ritsert C Jansen, GBIC Groningen Bioinformatics Centre, University of Groningen, Nijenborgh 7, 9747 AG Groningen, The Netherlands

Henk WM Hilhorst, Wageningen Seed Lab, Laboratory of Plant Physiology, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands

* Equally contributed to the presented work and should be considered first authors.
This work was supported by the Technology Foundation STW, the Applied Science Division of NWO (RVLJ, LAJW, WL); the Centre for BioSystems Genomics (CBSG) and the Netherlands Consortium of Systems Biology (NCSB), both of which are part of the Netherlands Genomics Initiative / Netherlands Organisation for Scientific Research (DA); the EU 7th Framework program under the Research Project PANACEA [222936] (YL, RJ).

Corresponding author: Wilco Ligterink, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands. wilco.ligterink@wur.nl
Abstract

A complex phenotype such as seed germination is the resultant of several genetic and environmental cues and requires the concerted action of many genes. The use of well-structured recombinant inbred lines in combination with omics analysis can help to disentangle the genetic basis of such quantitative traits. This so-called genetical genomics approach can effectively capture both genetic (G) and epistatic interactions (G:G). However, to understand how the environment interacts with genomic encoded information (G:E) a better understanding of the perception and processing of environmental signals is needed. In a classical genetical genomics setup this requires replication of the whole experiment in different environmental conditions. A novel generalized setup overcomes this limitation and includes environmental perturbation within a single experimental design. We developed a dedicated QTL mapping procedure to implement this approach and used existing phenotypical data to demonstrate its power. Additionally, we studied the genetic regulation of primary metabolism in dry and imbibed Arabidopsis seeds. Many changes were observed in the metabolome which are both under environmental and genetic control and their interactions. This concept offers unique reduction of experimental load with minimal compromise of statistical power and is of great potential in the field of systems genetics which requires a broad understanding of both plasticity and dynamic regulation.
Introduction

The use of natural variation to disentangle the genetic mechanisms underlying phenotypic differences has been very successful both in crop plants and in the model plant Arabidopsis thaliana (Alonso-Blanco et al., 2009). Most of the variation within wild or domesticated plant species is of quantitative nature determined by genetic polymorphisms at multiple loci. Such quantitative trait loci (QTL) can be analyzed efficiently using experimental mapping populations like recombinant inbred lines (RILs) derived from directed crosses. Nowadays, many well-structured RIL populations are available, often accompanied with detailed studies of phenotypic variation (Mitchell-Olds and Schmitt, 2006). The complexity of quantitative traits is further determined by the interactions between genomic loci, i.e. epistasis (G:G), and between the genotype and the environment (G:E). While epistasis can be effectively identified in QTL analyses, albeit with lower power than main effects, the detection of G:E interactions requires experimentation in multiple conditions of interest. Because of the large population sizes often needed to obtain sufficient statistical power for QTL detection, G:E interactions are usually ignored in experimental setups. However, a better understanding of the perception and processing of environmental signals is greatly needed, since interactions provide important insights in adaptation mechanisms and evolutionary constraints such as balancing and disruptive selection. To obtain a more detailed view of the molecular mechanisms underlying phenotypic variation, genetical genomics studies, in which molecular traits are genetically analyzed, have been successfully applied to enhance a directed strategy to identify causal relationships (Kliebenstein et al., 2006; Keurentjes et al., 2007; Van Leeuwen et al., 2007; Wentzell et al., 2007; West et al., 2007; Rowe et al., 2008). The observed phenotype is often the resultant of a functional cascade of gene transcription followed by protein translation and modification which finally leads to a highly dynamic metabolome underlying emergent properties (Kooke and Keurentjes, 2011). With the technological advances made in genomic analytical platforms, such as transcriptomics, proteomics, and metabolomics, the large-scale, high-throughput analyses needed for quantitative genetic approaches have become feasible (Jansen and Nap, 2001; Keurentjes et al., 2008). To incorporate developmental and environmental perturbation in the often expensive and laborious omic analyses, an alternative experimental setup, coined generalized genetical genomics (GGG), using
balanced fractions of a RIL population has been proposed (Li et al., 2008). It provides a cost-effective experimental setup for hypothesis generating research in multiple environments. Such an approach aims at the creation of sub-populations of RILs, one for each environment to be tested, with an optimal distribution of parental alleles over all available markers (Li et al., 2009). When these sub-populations are subjected to environmental perturbation the emerging phenotypes can be explained by several sources of variation: 1) genetic variation (G) 2) environmental variation (E) and 3) genetic x environmental variation (G:E). Whenever the resulting phenotype is not or only mildly affected by environmental interactions (G:E), the analysis of the different sub-populations can be combined gaining the full power of a complete population. However, when a trait shows strong G:E interaction, e.g. those that only express genetic variation in specific environments, the power to detect QTL is dependent on those sub-populations expressing the genetic variation. Although G:E interactions have been detected previously in genetical genomics studies for expression (Li et al., 2006; Smith and Kruglyak, 2008; Gerrits et al., 2009; Yeung et al., 2011) and metabolite content (Zhu et al., 2012) by analyzing all lines in a population under different environments, the GGG concept offers an effective way of studying a combination of genetic and experimental perturbations and is of great potential in the field of systems genetics in which a broad understanding of both plasticity and dynamics is required (Li et al., 2008). The fundamental basis of the experimental design and data analysis using a full model \( Y = E + G + G:E + e \) is generally valid and frequently used (Churchill, 2002; Li et al., 2006; Gerrits et al., 2009). As a proof of principle we present experimental data on the genetic regulation of primary metabolism in dry and imbibed *Arabidopsis* seeds using a GGG design and discuss the application and implications of such a strategy.

Plants are extremely rich in biochemical compounds and major roles in plant development, adaptation and defense have been identified for biosynthesis pathways and their products (Binder, 2010). The biosynthetic pathways of primary metabolites are well-studied and often well-conserved between different taxa (Peregrin-Alvarez et al., 2009). Nonetheless, quantitative variation for many of these compounds can be observed between natural variants which might be reflected in their different growth characteristics. The analysis of single gene mutants, for example, has unraveled many key components in biochemical pathways and has demonstrated their role in
phenotypic traits (Fiehn et al., 2000). In Arabidopsis, genetic variation for many of its metabolic compounds has been observed (Kliebenstein et al., 2001; Keurentjes et al., 2006; Rowe et al., 2008), but GxE interactions were ignored in these studies and only addressed by Chan and co-workers (2011). Metabolic profiling at different growth stages has further revealed important fluxes that regulate plant development and adaptation (de Oliveira Dal'Molin et al., 2010). Using the accumulated historical mutations that occur in natural variants in combination with metabolic profiling in a generalized design offers the unique possibility of identifying genetic effects over a series of developmental stages. Here we report on the interaction of four different physiological environments, i.e. developmental stages, in dry and imbibed seeds with two founder genotypes in a RIL population. To detect the majority of the most prominent primary metabolites we used gas chromatography-mass spectrometry (GC-MS) of polar extracts (Roessner et al., 2000; Lisec et al., 2008). These include essential metabolites such as sugars, amino acids, and organic acids, which are key compounds in reserve storage and catabolism, growth and energy metabolism.

The switch from a dry seed, which is equipped for optimal survival and storage of reserves, towards an imbibed seed, in which energy needed for germination is released and which prepares for autotrophic production is remarkable. Reserves that have been stored during seed maturation are degraded and remobilized during germination (Bewley, 1997; Shu et al., 2008), a process that is heavily influenced by the capacity of C/N partitioning of a maturing seed (Dowdle et al., 2007). Arabidopsis mutants affected in their oil reserve content or its mobilization show delayed, but not full inhibition of germination (Kinnersley and Turano, 2000; Bouche and Fromm, 2004; Shu et al., 2008; Kelly et al., 2011). This suggests an additional metabolic switch that occurs during seed desiccation after seed maturation, involving a change from accumulation of oil and storage proteins to the synthesis of free amino acids, sugars, fatty acids and their degradation products functioning to prepare for rapid metabolic recovery during imbibition (Fait et al., 2006; Angelovici et al., 2010).

Imbibition of mature seeds specifically shows reduction of the metabolites that accumulate during the desiccation period. Upon germination, an increase of many metabolites, including amino acids, sugars and organic acids, can be observed again, which reflects the increase of autotrophic activity (Fait et al., 2006). Profiling the primary metabolome over different developmental stages in a mapping population
is therefore expected to reveal the dynamics of genetic regulation of many of these important processes. We will demonstrate here that much of the observed variation in biochemical profiles can be attributed to genotype-by-environment interactions which can be effectively identified in a generalized genetical genomics approach.

Results and Discussion

Experimental design

Previous studies which focused on the comparative analysis of developmental and metabolic variation suggest a link between central metabolism and plant physiology, but genetic co-regulation is not frequently observed (Keurentjes et al., 2006; Meyer et al., 2007). That said, in several studies in Arabidopsis a major metabolite QTL cluster is associated with the ERECTA locus, representing a strong regulator of development which is known for its pleiotropic effects (Fu et al., 2009). To circumvent this strong bias we used two natural variants, Bayreuth-0 (Bay-0) and Shahdara (Sha), which are not polymorphic for the ERECTA locus. The Bay-0 x Sha RIL population (Loudet et al., 2002) has previously been shown to contain genetic variation for seed germination (Joosen et al., 2012) and other physiological traits (Loudet et al., 2003; Barriere et al., 2005; Loudet et al., 2005; Diaz et al., 2006; Reymond et al., 2006; Loudet et al., 2008; Meng et al., 2008), anion strength (Loudet et al., 2003), carbohydrate content (Calenge et al., 2006), gene expression (West et al., 2007) and primary (Rowe et al., 2008) and secondary metabolite levels (Wentzell et al., 2007).

Powerful mapping of genetic variation in a RIL population is dependent on the size of the population, the level of recombination and on an evenly genome-wide distribution of the parental alleles. In the present study, a core set of the Bay-0 x Sha RIL population (Loudet et al., 2002) consisting of 165 lines and optimized for the aforementioned factors was used. This core population was divided in four sub-populations optimized for the distribution of parental alleles using the R-package DesignGG, aiming at the most accurate estimate of genetic and G:E effects (Li et al., 2009, Supplemental Fig. 1).
Comparison of different designs using classic phenotypes

Standard QTL mapping procedures can efficiently capture genetic variation and epistasis, but do not take environmental perturbation into consideration. Appropriate modeling of the genetic variance-covariance (VCOV) in the data is of great importance when combining information from different environments in QTL analysis (Churchill, 2002). Linear models are particularly well suited for this. Here environmental differences are incorporated as an additional variable in a generalized design (GGG design). To enable mapping of the observed trait variation and taking the four developmental stages into consideration an R-script was developed which use functions and data structures from the R/qtl package (Broman et al., 2003; Arends et al., 2010, Supplemental file 3). The R-script uses a linear model to calculate the likelihood of genotype to phenotype linkage for each marker with the following formula:

\[ y_i = \beta_0 + \beta_1 e_i + \beta_2 g_i + \beta_3 g_i e_i + \epsilon_i \]

Where \( y_i \) is the \( i \)th observation of the studied phenotype, variable \( g_i \) is the genotype, \( e_i \) is a vector with seed conditions, and \( g_i e_i \) the interaction term. The values \( \beta_j \) represent parameters to be estimated, and \( \epsilon_i \) is the error term. The simplified description \( Y = E + G + G:E + \epsilon \) of this linear model will be used henceforward. Separate likelihood estimates (\(-\log_{10}\) Probability, henceforth LOD scores) are generated for the environmental (E), genetic (G) and genetic x environmental (G:E) effects.

To validate the use of a GGG design, we studied the genetic (G) and the interacting effects between G and E (GxE) on phenotypes in four different environmental conditions (E). These phenotypes were obtained by studying different germination parameters under different environmental conditions (Joosen et al., 2012). In total we compared the power of different designs by performing QTL analysis for 96 classic phenotypes under 4 different environments (Joosen et al., 2012) (table 1). Furthermore, we also investigate the interacting effect between genotype and environment. The full model mapping \( Y = E + G + G:E + \epsilon \) was applied to a full block design, random design and GGG design. Single marker mapping \( Y = G + \epsilon \) was applied to a single block design. The number of detected
QTL and interacting QTL (FDR = 0.05, based on >10000 runs permutation) with the different designs are shown in table 1. In the full block design all samples were allocated to the four conditions. Obviously, this is the most expensive way of performing the experiment as the required resources and effort are quadrupled (4 x N). As a consequence of the size of the experiment, the power of detecting genetic effects is the best for this design. Unfortunately, we cannot afford such expensive experiments in many situations due to limited resources and time. The single block design only focuses on one of the four conditions, as in most published genetical genomics studies to date. In this way the samples size for the selected condition is N and we will have equal power as in the full block design for detecting the genetic effects for this particular condition. Clearly, this design will miss the information from the other three conditions and interacting effects between genetic and environmental factors cannot be investigated. In order to study both genetic and interacting effects with a limited budget, the random and the GGG design allocate the N different samples to the four environments evenly, measuring N/4 samples in each condition. Although the possibility to detect genetic effects is only slightly better for the GGG design, the detection of interacting QTL is clearly improved in the GGG design as compared to the random design. These results show that the optimal allocation of samples in the GGG design clearly improves the ability to detect both genetic and interacting effects and that the GGG design results in the maximization of detected variation in relation to the necessary resources with only a minimal compromise of statistical power as compared to the full block design.

Metabolic analyses

To study the metabolic status of Arabidopsis seeds during germination, four biologically important developmental stages of seed germination with expected variation in metabolite levels to different extent were selected. The first two stages, being freshly harvested primary dormant (PD) and after-ripened (AR) non-dormant dry seeds, respectively, are expected to comprise a very similar metabolome as most, if not all, metabolic fluxes are arrested in the dry seed. The oil rich (~40%) Arabidopsis seeds (Hobbs et al., 2004) typically desiccate to moisture contents below 5% which results in an arrest of all enzymatic reactions due to the lack of free water. The other two stages represented early imbibition of seeds, imbibed for 6 hours (6H),
and seeds at radical protrusion (RP), respectively. Full rehydration of dry seeds typically completes in less than 2 hours and although developmental differences are not yet expected, many metabolic processes will have started after 6 hours of imbibition (Nakabayashi et al., 2005; Howell et al., 2009). Radicle protrusion marks the end-point of germination *sensu stricto* and is known to be accompanied by a major switch of both the transcriptome and metabolome (Nakabayashi et al., 2005; Fait et al., 2006). These four developmental stages are anticipated to vary to different degrees in their metabolic profiles, hardly any difference between dry seed samples, some differences between dry and imbibed seeds and very pronounced differences between dry seeds and seeds at radicle protrusion.

To determine the metabolic status of genetic variants in these different developmental stages, all individuals in the four sub-populations and their parental accessions were subjected to GC-TOF-MS. Each sample consists of the polar fraction of a methanol extract of a bulk of approximately 700-1000 seeds (20 mg). Samples were analyzed in random order and interspersed with pooled sample controls to control for experimental errors. The metabolic profiling of the segregating RILs was performed and the use of segregation population provides an intrinsic replication for each genotypic marker (Jansen and Nap, 2001). In total 7537 mass peaks were detected, representing 161 metabolites according to centrotyping based on retention time and correlation structure (Tikunov et al., 2011). In total 63 metabolites could be annotated using an in-house constructed library and a publicly available mass spectra library (Schauer et al., 2005, Supplemental file 1).

The parental accessions Bay-0 and Sha were measured in duplicate for all four developmental stages allowing us to model the influence of condition and accession using a multi-factor univariate analysis of variance (ANOVA).

\[ y_i = \beta_0 + \beta_1 \text{condition}_i + \beta_2 \text{accession}_i + \epsilon_i \]

Analysis of variance for the parental samples identified 108 metabolites showing significant variation (FDR < 0.05) between developmental stages (E) and 85 showing variation between the parents (G) with an overlap of 54 metabolites showing variation between both variables in an interactive way (G:E) (Supplemental file 2). For 37 metabolites no significant variation was detected between the parental
accessions or in any of the developmental stages. A self-organizing map (SOM), created from the metabolites showing significant variation between the parents, groups different metabolites according to their accumulation pattern over different genotypes and developmental stages (Fig. 1). Clearly different patterns of variation can be observed, namely genetic in panel A and H; environmental in panel C and D; genetic + environmental in panel B and G and genetic × environmental in panel E and F, illustrating the complex regulation of metabolic processes and the need for sophisticated analysis methods, like PCA or Multiple QTL mapping (Arends et al., 2010).

Because metabolite levels are varying between both parents and between the chosen seed germination stages, a segregation of metabolic accumulation can be expected in the RIL population of 164 lines. A principle component analysis of the metabolic profiles, revealing the internal structure in the data, shows that the first component clearly separates 6-hour imbibed seeds and seeds at radicle protrusion from both primary dormant and after-ripened seeds, explaining 37% of the total variation (Supplemental Fig. 2). This confirms the large metabolic changes accompanying the transition from dry arrested seeds to the imbibed and germinating developmental stages. As expected, no obvious differences could be detected between the metabolomes of primary dormant and after-ripened dry seeds. The second component, explaining 11% of the total variation, separates the parental accessions, indicating that this component explains a lot of the genetic variation in metabolic profiles. These results demonstrate that Bay-0 and Sha possess genetic variation for the accumulation of primary metabolites which segregates in their recombinant offspring and which is strongly influenced by the developmental stage used for profiling.

Transgressive segregation was visualized by comparing parental and RIL metabolite level distributions (Supplemental Fig. 3). Some positive and negative transgression is observed for most of the metabolites in which the metabolite accumulation in a RIL is respectively higher or lower compared to the respectively highest or lowest parent. In addition, 15 metabolites were detected in RILs which were not present in either parent. This suggests that new allele combinations in the RIL population resulted in enhanced accumulation or even novel formation of metabolites, although it could also be that those metabolites were missed in the...
parents because of the limits of the technology and methodology used in the current study.

**Genetic mapping of metabolites in a generalized genetical genomics design**

In the experimental setup of this study, the environmental variation is defined as variation observed between the four developmental stages (PD, AR, 6H and RP). Significance thresholds, determined by permutation analysis (n=1000, p<0.01) for each metabolite, ranged from LOD 3.43 to LOD 3.50 and was stringently set to LOD 4 for all analyses. Mapping resulted in 120 significant QTLs in the genetic (G) component for 83 metabolites and 31 genetic x environmental (G:E) QTLs for 27 metabolites, ranging from one to four QTLs per metabolite. Thirteen of the G:E QTLs are significant in the G component as well. For 66 metabolites no significant QTL was detected. Clustered heat maps for both the G and the G:E QTL profiles were created (supplemental Fig. 4 and 5).

To test the performance of the generalized mapping procedure, QTLs detected in individual environments (using the linear model \( y_i = \beta_0 + \beta_1 g_i + \varepsilon_i \), henceforth \( Y = G + \varepsilon \) ) were compared to QTLs detected in the combined mapping approach (using the linear model \( Y = E + G + G:E + \varepsilon \) ) (Fig. 2, supplemental table 1). QTLs were binned in upper or lower chromosome arms to reduce the effects of small positional shifts. Results were plotted in a network with nodes representing QTLs connected with edges to nodes representing the mapping populations in which they were detected (Fig. 2). QTLs are grouped in three panels according to their detection in the different mapping procedures. The middle panel shows 73 QTLs that were detected in both the \( Y = E + G + G:E + \varepsilon \) model and in one or more single environment mappings using the \( Y = G + \varepsilon \) model. This shows that most of the genetic variation present in the single environments can effectively be captured by using the generalized model. The presence of 60 QTLs that were only significantly detected in the \( Y = E + G + G:E + \varepsilon \) model (right panel) shows the combined power of the generalized approach and the usage of more genotypes. These QTLs are not detected in the single environment mapping in which only 41 individuals were used. Combining all data across all environments in the linear model increases power to detect QTLs, but it should be noted that there are also 20 minor QTLs (left panel) which are only significant in the single environment mapping with the \( Y = G + \varepsilon \).
model. These QTLs are not detected in the \( Y = E + G + G:E + \varepsilon \) model. This can be explained by two factors: 1) environments in which the genetic variation is not expressed introduce noise in the experimental data and thereby decrease mapping power, and 2) deviations from a balanced allele distribution in the different subpopulations can introduce some stochasticity around the threshold level, although this is not the case in our data.

Importantly, all major to moderate effect size QTLs could be detected using the generalized model even when these QTLs were not detected in the separate environment models. Although it is difficult to compare power with the latter models, because population sizes differ, the generalized design efficiently identifies all relevant QTLs which were detected by the four separate models and in addition it detects G:E interactions. In a general exploratory study, the reduction in experimental burden therefore amply outweighs the incidental failure to detect the limited number of small-effect QTLs. The application of a GGG design can thus be an important advancement in evolutionary and ecological studies assessing the contribution of genetic and environmental effects to natural variation in life history traits.

For breeding purposes the allelic effect size is an important measure and differentiation of the environment in which the allelic effect is expressed can be very useful. In the generalized setup the allelic effect size of those metabolites with significant QTLs is separated per environment (Supplemental files 4 and 5). For every QTL which is consistently detected in all four conditions, a LOD score for genetic effect (x axis in Fig.3) is obtained from full model mapping. For these QTLs, normalized allelic effect sizes are calculated by Z-score transformations for each environment (y axis in Fig.3). QTLs detected in the G component of the linear model (Fig. 3A) show an expected linear relationship between LOD score and effect size in all measured environments. This correlation is much weaker for QTLs detected in the G:E component of the linear model (Fig. 3B), because the genetic variation is not expressed in all environments. QTLs of metabolites with strong G:E interaction, therefore, display larger effect sizes in fewer environments compared to G component QTLs of similar significance levels.

Clearly, the choice of environments used in such study is crucial (Li et al, 2008). Limited power can be expected when environments vary too much and no overlapping genetic variation is present and contrarily there is hardly additive value of
the design when using very similar environments. In this study we carefully selected
four biologically relevant developmental stages of seed germination with expected
variation in metabolite levels to different extent and consider them as an
environmental factor in the follow-up statistical analysis. The selected developmental
stages starts from primary dormant dry seeds to seeds at the point of radicle
protrusion. The first two stages, being freshly harvested primary dormant (PD) and
after-ripened (AR) non-dormant dry seeds, respectively, are expected to comprise a
very similar metabolome as most, if not all, metabolic fluxes are arrested in the dry
seed. The other two stages represent seeds imbibed for 6 hours (6H), and seeds at
radical protrusion (RP), respectively. Different levels of environmental variation were
obtained and could be mapped by the genetic (G) and/or genetic x environment (G:E)
component of the linear model.

Genetic regulation of metabolic traits

One of the most rewarding benefits of the generalized approach is the
possibility to analyze metabolic fluxes over different environments or developmental
stages in addition to the effect of genetic variation. The acquired information of both
sources of variation can be effectively displayed in so-called flash cards in which line
graphs illustrate the genetic and environmental effect and detected QTLs are plotted
in heat bars (Fig 4; Supplemental Fig. 6). The individual components of the linear
model $Y = E + G + G:E + \varepsilon$ provide the valuable measures for the various sources of
variation. For example lysine content strongly increases in germinating seeds,
indicated by a significant LOD score of 16.1 for the environmental effect, but no
genetic variation for lysine could be detected (Fig 4A). For this metabolite genetic
variants vary indistinguishable from each other over different environments. In
contrast, fumaric acid shows little variation between the developmental stages (LOD
0.6), but displays strong genetic variation explained by a highly significant QTL (LOD
6.5) for the genetic effect at the center of chromosome II. Higher levels for fumaric
acid are detected in all developmental stages for those lines harbouring the Bay-0
allele (Fig 4B). An example of the additive effect of environmental and genetic factors
is the decrease in levels of malic acid in imbibed seeds. Here a strong environmental
effect (LOD 13.2) is accompanied with an additional genetic effect explained by a G
QTL (LOD 6.9) at the bottom of chromosome I. Note that the genetic effect here is
similar in all environments (Fig 4C). This is not the case for gluconic acid which levels
are strongly affected by the interaction between the genotype and the environment. A
strong G:E QTL (LOD 10) is detected at the top of chromosome IV. The Sha allele at
this position causes higher levels of gluconic acid in dry seeds, but not in imbibed
seeds (Fig 4D). This strong negative environmental effect (LOD 6.6) is also
responsible for the apparent directional shift of the G:E QTL effect.

Similar to the self-organizing maps in figure 1 flashcards can be instrumental
in the identification of metabolic relationships with the added value of genetic
regulatory information. This is illustrated by integrating flashcards of all metabolites
that were identified in this study with a general Arabidopsis metabolic pathway
diagram (http://www.KEGG.jp, Supplemental Fig. 7). For instance, several pathways
in carbohydrate metabolism, such as the biosynthesis routes for galactose, pentose
phosphate, starch/sucrose and amino and nucleotide sugars, are highly
interconnected and are therefore subject to co-regulation mechanisms. A number of
compounds involved in different subparts of the carbohydrate network module (e.g.
glucose-6-phosphate, maltose, mannose, glucuronic and gluconic acid) indeed share
a strong QTL at the top of chromosome IV. This suggests that the observed variation
for these compounds has a single genetic basis, possibly affecting competition for a
general precursor or directing feedback loops. In addition many of these compounds
show strong positive or negative correlation due to environmental control. Genetic co-
regulation was also observed for amino acid metabolism. Amino acids are substrate
for the synthesis of aminoacyl-tRNAs which in turn are essential substrates for
translation (Sheppard et al., 2008). A single G:E QTL at the bottom of chromosome I
was detected for eight amino acids explaining a large part of the observed genetic
variation. The joined analysis of environmentally and genetically induced variation in
metabolic profiles can thus identify causal relationships between different modular
parts of metabolic networks and associate these connections with relevant biological
processes.

Regulatory hotspots and physiological co-regulation

As noted, the accumulation of several metabolites maps to identical positions
suggesting that these might be regulated by a common genetic factor. Although co-
locating QTLs can be the result of independent closely linked genetic factors, such coinciding QTLs are expected to occur more or less randomly by chance. Any deviation from expected frequency distributions along the genome thus hints at genetic co-regulation (Breitling et al., 2008). When plotted against their genomic position eight of such suggestive QTL hotspots can be seen (Fig. 5) of which the two major ones (Chromosome IV-MSAT4.8 and Chromosome V-NGA139) co-locate with previously identified hotspots for metabolic regulation (Kliebenstein et al., 2001; Keurentjes et al., 2006; Wentzell et al., 2007; Rowe et al., 2008). Interestingly, both these loci have been shown to play a role in glucosinolate biosynthesis. The AOP locus at chromosome IV regulates side chain modification while the MAM locus at chromosome V determines chain elongation, but these compounds are not targeted for in GC-MS analysis which predominantly detects primary metabolites. As for many glucosinolates, for some metabolites, including GABA and maltose, QTLs were detected at both positions. In other cases a single QTL was detected at chromosome IV or V, e.g. glucose-6-phosphate and tyrosine, respectively. Although the identified primary metabolites are not directly connected with the glucosinolate biosynthesis pathway such associations have been reported before (Rowe et al., 2008). These results might suggest alternative functions for AOP and MAM or a role in resource competition and allocation in central metabolism. This suggestion is further supported by the fact that these loci link to flowering time and the circadian clock regulation in the Bay-0 x Sha population (Chan et al., 2011). It also cannot be ruled out that other genes overlapping the AOP or MAM regions are causal for the observed variation.

Since many metabolites appear to be co-regulated, the strong impact of some loci on central metabolism might also exert its effect on physiological traits. Recently, the genetic landscape of seed germination in the same population has been described for which seed germination parameters were acquired under a wide range of environmental conditions (Joosen et al., 2012). A comparison between variation in germination characteristics and metabolite levels might reveal compounds involved in the process of germination. Although no clear co-location of hotspots for germination and metabolite QTLs could be observed, incidental coincidence between isolated QTLs of both types of traits did occur. For instance, genetic variation for seed size co-locates with a large metabolic QTL cluster on the lower arm of chromosome I (~75 cM). This cluster contains many QTLs for amino acids, but also for components of
the TCA cycle (e.g. fumarate and malate). In plants, leucine, isoleucine and valine, can be broken down and the end products of their catabolic pathways enter the TCA cycle to generate energy. It has been shown that these amino acids promote their own degradation, but only during seed germination, senescence, or under sugar starvation (Binder, 2010). This suggests that the degradation pathways provide alternative carbon sources for the plant in extreme conditions. In addition, branched-chain amino acids and their derived alpha-keto acids are cytotoxic and preventing accumulation through degradation may be an important detoxification mechanism (Fujiki et al., 2000). Higher levels of both fumarate and malate, as a result of the degradation of a surplus of amino acids, might thus be indicative for larger seed sizes. A second QTL for seed size on chromosome V co-locates with a QTL of opposite effect for GABA accumulation. Interestingly, Bay-0 alleles at both QTLs confer larger seed size, suggesting that there was selection pressure for large seed size in the environment where Bay-0 was collected, as was also observed in a different population (Alonso-Blanco et al., 1999). However, where levels of fumarate and malate are increased in larger seeds, the accumulation of GABA is decreased. GABA is known to be involved in a range of cellular processes (Palanivelu et al., 2003) and is rapidly accumulated in response to biotic and abiotic stresses (Kinnersley and Turano, 2000). It has been postulated that it has roles in herbivore deterrence, pH and redox regulation, energy production and maintenance of carbon/nitrogen (C/N) balance (Bouche and Fromm, 2004). In a recent study, GABA levels in seeds were shown to increase by expressing glutamate decarboxylase (GAD) under a seed maturation-specific phaseolin promoter (Fait et al., 2011). In accordance with our findings this resulted in smaller seed size and reduced seed vigor in T3 plants. No opposite seed size effect could be detected at a GABA QTL with increased levels due to the Bay-0 allele at the top of chromosome four, but colocating genetic variation for germination on ABA, heat sensitivity and dormancy was observed at this position. These cases illustrate the power of joined genetic analyses of metabolic and physiological traits for generation of hypotheses that can help in the functional annotation of plant metabolites and their possible role in the regulation of important physiological processes.

Confirmation of mQTLs
To independently confirm the effect of a single locus it must be isolated and tested in an isogenic background. Several methods can be followed to perform such an independent confirmation of QTLs. A powerful approach is the use of residual heterozygosity in early generations of RILs. The Bay-0xSha RIL population (420 lines in total) was genotyped at F6 in which approximately 97% homozygosity is reached in each line. This resulted in the presence of residual heterozygosity in at least a single RIL at almost all genome positions. Those heterozygous regions are segregating in a Mendelian fashion in the next generation and can be used to confirm QTL positions, as it provides a possibility to study both parental alleles at the locus of interest in an otherwise homozygous background (Tuinstra et al., 1997). In a heterogeneous inbred family (HIF) those heterozygous regions are fixed and two separate lines containing the alleles of both parents respectively are maintained.

HIF312 and HIF214 are segregating for regions at the top of chromosome IV and V (Fig. 6A), respectively, and cover the region in which the two major metabolite hotspots were detected. After ripened dry seeds were used to profile the HIFs for metabolic content because many of the QTLs detected in this region showed a large-effect size at the dry seed stages. Significant differences between parental alleles using 4 replicates were defined by a two-tailed t-test ($p<0.05$). In total 34 out of 64 QTLs could be confirmed using this approach (Supplemental Fig. 8). For maltose for instance, two QTLs with opposite direction were found (Fig. 6B) which both could be confirmed using the two distinct HIFs (Fig. 6C). In a number of cases a HIF effect was observed that was not detected significantly in the RIL population (e.g. Digalactosylglycerol). This might be the result from the higher power in near isogenic lines due to the absence of epistatic interactions (Keurentjes et al., 2007). Nonetheless, a substantial number of QTLs could not be confirmed by the HIF lines. The enrichment for small-effect QTLs in the unconfirmed class suggests that four replicates generate insufficient power to identify significant differences for these metabolites in the HIF experiments, although we cannot rule out that they are false positives from the QTL analysis. Furthermore, QTLs depending on epistatic interactions cannot be detected in some near isogenic lines. In addition, a number of QTL support intervals are broader than the region covered by the HIF and thus the causal genetic polymorphism within the QTL interval, but outside the region covered by the HIF, would have been missed.
The analyses of the HIF lines indicate that most of the large-effect QTLs can be accurately detected using a generalized genomics approach. Although an underestimation of small-effect QTLs can be expected this is largely compensated by the higher power of detecting genetic and environmental interactions.

Conclusions

The use of natural variation is a valuable tool to dissect the genetics of complex traits and the addition of powerful ‘omics’ analysis provides a great resource to disentangle molecular mechanisms. However, the expensive nature of many ‘omics’ experiments limits researchers to deploy perturbation of either environment or development. New strategies are needed to enable the switch from genetical genomics to system genetics. Here we have reported on a strategy to divide a RIL population in well-defined sub-populations and to use those to perturb the environment or developmental stage. To this end a novel R-script has been created to enable QTL mapping using a linear model that includes the possibility to account for genetic and environmental variation. This R-script is fast enough to analyze hundreds to thousands of traits and creates possibilities to extend the generalized genetical genomics strategy to whole genome gene expression analysis by either microarray or next generation sequence approaches (Joosen et al., 2009; Ligterink et al., 2012).

Efficient QTL mapping is strongly dependent on the population size and recombination frequency. Keurentjes and co-workers (2007) studied the effect of the population size and showed a linear relationship between the number of individuals used for mapping and the smallest detectable genetic effect. In this light it might seem undesirable to split a RIL population in smaller sub-populations. This is true when genetic variation is only detectable in a single unique environment or developmental stage leading to a strong genetic x environment interaction. More often, variation is subject to the environment without a complete abolishment of the genetic variation. In those cases the environmental effects can be normalized and the power of detecting a QTL is increased to the total number of lines used in the different sub-populations. The availability of a genome wide set of heterogeneous inbred family (HIF) lines of the Bay-0 x Sha RIL population provides a solid and fast way to confirm QTLs. By using this approach we tested two of the observed QTL
hotspots and were able to confirm many of the detected QTLs. When resources are limited this can be regarded as a good alternative for replicating the whole experiment for e.g. different growth seasons.

Many studies have shown the highly dynamic nature of molecular mechanisms leading towards seed germination (e.g. reviewed in Catusse et al. (2008), Daszkowska-Golec (2011) and Weitbrecht et al. (2011)). Performing expensive genetical genomic experiments without any perturbation of the environment will therefore always raise questions about the possible extrapolation of the results when slightly different conditions are used. Information about the flux of a metabolite within a range of developmental stages or within a range of environments allows a much more precise interpretation of the molecular effects. By using the generalized strategy we showed that it is possible to deduct the metabolic fluxes (Fig. 4). This extra level of information is a very valuable addition and helps to interpret the effect of genetic variation in the context of a dynamic and constantly changing metabolome.

Metabolite hotspots can reveal important loci involved in major metabolic pathway differences between two natural variants. In several studies the detected ‘omics’ hotspots did not co-locate more than expected by chance with phenotypic hotspots (Keurentjes et al., 2006; Meyer et al., 2007). However, in this study we detected some co-locating QTLs which might be explained by the narrow developmental window in which both metabolite and phenotypic QTLs (Joosen et al., 2012) were gathered. We detected overlapping QTLs for amino-acid synthesis, TCA cycle compounds and seed size at the bottom of chromosome I and also co-location between QTLs for GABA, seed size and germination under stress conditions at the top of chromosome 5 (Joosen et al., 2012). These co-locating QTLs are interesting leads for further research which is necessary to elucidate the true causal molecular mechanisms.

In conclusion, in the era of large systems genetics initiatives, we propose to consider the use of a generalized design for genetical genomics studies. The simultaneous acquisition of both genetic variation and developmental fluxes is a cost effective approach enabling a much better understanding of the processes involved. We see great potential in further exploration of the generalized design for transcriptome or other ‘omics’ related studies.
Material and methods

Plant material

Seeds from the core population (165 lines) of the *Arabidopsis* Bay-0 x Sha recombinant inbred line population (Loudet et al., 2002) and heterogeneous inbred family (HIF) lines were obtained from the Versailles Biological Resource Centre for *Arabidopsis* (http://dbsgap.versailles.inra.fr/vnat). The population is mapped with 69 markers with an average distance between the markers of 6.1 cM (Loudet et al., 2002). Maternal plants were grown in a fully randomized setup and seeds from 4-7 plants/RIL were bulk harvested. Plants were grown on 4x4 cm rockwool plugs (MM40/40, Grodan B.V.) and watered with 1 g/l Hyponex fertilizer (NPK=7:6:19, http://www.hyponex.co.jp) in a climate chamber (20°C day, 18°C night) with 16 hours of light (35 W/m²) at a relative humidity of 70%. Seeds were either stored at -80°C 1 week after harvest (primary dormant; PD) or after-ripened at room temperature and ambient relative humidity until maximum germination potential after 5 d of imbibition was reached (after-ripened; AR). After-ripened seeds were imbibed on water saturated filter paper at 20°C for 6H and quickly transferred to a dry filter paper for 1 minute to remove excess of water (6 hours imbibed seeds; 6H) Manual selection with help of a binocular was carried out to harvest seeds with the radicle at the point of protrusion (radicle protrusion; RP). Three radicle protrusion lines failed the metabolite analysis and were replaced by dry primary dormant samples.

Metabolite analysis

The metabolite extraction was performed based on a previously described method (Roessner et al., 2000) with some modifications. Seeds (20 mg) were homogenized using a micro dismembrator (Sartorius) in 2 ml tubes with 2 iron balls (2,5 mm), precooled in liquid nitrogen. 700 µl methanol/chloroform (4:3) was added together with the standard (0.2 mg/ml ribitol) and mixed thoroughly. After 10 minutes of sonication 200 µl MQ was added to the mixture followed by vortexing and centrifugation (5 min., 13500 rpm). The methanol phase was collected in a glass vial. 500 µl methanol/chloroform was added to the remaining organic phase and kept on ice for 10 min. 200 µl MQ was added followed by vortexing and centrifugation (5 min., 13500 rpm). Again the methanol phase was collected and mixed with the other
collected phase. 100 µl was dried overnight using a speedvac (35°C Savant SPD121).

A GC-TOF-MS method (Carreno-Quintero et al., 2012) was used with some minor modifications. Detector voltage was set at 1600V. Raw data was processed using the chromaTOF software 2.0 (Leco instruments) and further processed using the Metalign software (Lommen, 2009), to extract and align the mass signals. A signal to noise ratio of 2 was used. The output was further processed by the Metalign Output Transformer (METOT; Plant Research International, Wageningen) and mass signals that where present in less than 3 RIL’s where discarded. Centrotypes were created using the MSclust program (Tikunov et al., 2011). The mass spectra of these centrotypes were used for the identification by matching to an in-house constructed library and the NIST05 (National Institute of Standards and Technology, Gaithersburg, MD, USA; http://www.nist.gov/srd/mslist.htm) libraries. This identification is based on spectra similarity and comparison of retention indices calculated by using a 3\(^{\text{rd}}\) order polynomial function (Strehmel et al., 2008).
QTL mapping

Data was preprocessed using a log10 transformation and per phenotype outliers were removed after Z-transformation (Z-scores > 3). With the open source statistical package R (version 2.14.1) we fitted a basic linear model \( y = \beta_0 + \beta_1 g_i + \epsilon_i \) on the 4 conditions separately. This was followed by a combined mapping allowing for a developmental covariate and interaction term between the genetic marker and developmental stage \( y = \beta_0 + \beta_1 e_i + \beta_2 g_i + \beta_3 e_i g_i + \epsilon_i \). P-values from all mappings are transformed into LOD scores by taking the \(-\log_{10}\). Additionally, raw and normalized effects were calculated for each individual environment. Normalized effects were calculated by dividing the difference between the maximum and the minimum value for that trait by the mean effect at the marker. LOD significance was determined using permutations for the combined mapping of the 4 environments: a LOD score of 4 was found to be significant (Breitling et al., 2008). Supplemental file 3 contains the R script used for the data analysis.

Acknowledgements

We would like to thank Linus van der Plas for critical reading of the paper.

Supplemental data files

Supplementary Table 1: Overview of QTLs shared between different models based on 95% confidence intervals.

Supplemental file 1: Metabolite centrotype data. Including peak numbers, retention time, hit quality, probability and mass used for quantification.

Supplemental file 2: ANOVA results from metabolic profiling of the parental lines Bay-0 and Sha.

Supplemental file 3: R-script with original data files allowing re-analysis of all data provided in this paper.

Supplemental file 4: Summary of all detected metabolic G QTLs.

Supplemental file 5: Summary of all detected metabolic G:E QTLs.
Supplemental figure 1: Allele distribution within the Bay-0 x Sha RIL population and the 4 selected sub-populations. Blue indicates the percentage of lines with a Bay-0 allele for a certain marker and red the number of lines with a Sha allele.

Supplemental figure 2: Principal component analysis plot showing the first two principal components of the metabolite analysis in the Bay-0 x Sha RIL population. Colors indicate the developmental stage (red = primary dormant (PD); blue = after-ripened (AR); green = 6 hours imbibed (6H); orange = seeds at radicle protrusion (RP), parental lines are indicated by triangles (Sha) or squares (Bay-0).

Supplemental figure 3: Transgression plot. Graph with scaled metabolite levels per RIL and parental levels.

Supplemental figure 4: Clustered heat map from the genetic (G) component showing the LOD profiles of all metabolites. Columns indicate marker positions along the five chromosomes; rows indicate individual-trait LOD profiles. A false-color scale is used to indicate the QTL significance. Positive values (yellow and red) represent a larger effect on the metabolite content for the Sha allele, and negative values (blue and green) represent a larger effect on the metabolite content for the Bay-0 allele. Clustering on the left shows correlation between QTL profiles.

Supplemental figure 5: Clustered heat map from the genetic x environmental (G:E) component showing the LOD profiles of all metabolites. Columns indicate marker positions along the five chromosomes; rows indicate individual-trait LOD profiles. A false-color scale is used to indicate the QTL significance. Positive values (yellow and red) represent a larger effect of the treatment for the Sha allele, and negative values (blue and green) represent a larger effect of the treatment for the Bay-0 allele. Clustering on the left shows correlation between QTL profiles.

Supplemental figure 6: Flashcards of all identified metabolites. For full legend see figure 4.

Supplemental figure 7: KEGG metabolic pathway with flashcards overlay of the metabolites identified in this study.

Supplemental figure 8: Overview from HIF analysis with all metabolites with significant QTL confirmation.

Literature


Kelly AA, Quettier AL, Shaw E, Eastmond PJ (2011) Seed storage oil mobilization is important but not essential for germination or seedling establishment in Arabidopsis. Plant Physiology 157: 866-875


www.plantphysiol.org on July 14, 2017 - Published by Downloaded from Copyright © 2013 American Society of Plant Biologists. All rights reserved.
Ligterink W, Joosen RVL, Hilhorst HWM (2012) Unravelling the complex trait of seed quality: using natural variation through a combination of physiology, genetics and -omics technologies. Seed Science Research 22: S45-S52


related to high plant growth rate in Arabidopsis thaliana. Proc Natl Acad Sci U S A 104: 4759-4764


Tikunov YM, Laptenok S, Hall RD, Bovy AG, Vos CHd (2011) MSClust: a tool for unsupervised mass spectra extraction of chromatography-mass spectrometry ion-wise aligned data (online first). Metabolomics


Figure Legends

Figure 1: Self organizing map, grouping different metabolites according to their accumulation pattern over different genotypes and developmental stages of significantly variable metabolites (anova F pr. < 0.05) measured in the parental lines Bay-0 and Sha in four developmental stages. PD=Primary dormant, AR=After-ripened, 6H=6 hour imbibed, RP=seeds at radicle protrusion. Two independent biological replicates were measured for each combination of parent and developmental stage.

Figure 2: Comparison of QTLs detected within single environments (PD, AR, 6H and RP) by using the simple $Y=G+\varepsilon$ model with QTLs detected when combining environments via the full $Y=E+G+G:E+\varepsilon$ model. QTLs were binned to two regions per chromosome (i.e. top and bottom region). When comparing QTLs of a single trait from two models, they are considered as shared ones if QTLs fall in the same region. In total we found 73 QTLs shared between the two models, as shown in the middle ellipse. There are 20 and 60 QTLs that are only detected in the simple and full model, respectively. Nodes indicate metabolite QTLs and node size shows the degree of connectivity. Nodes are connected by edges which show the link between a QTL and a mapping population (single environments versus multiple environments). Separate nodes are created for the genetic (G) component and the genetic x environmental (G:E) component. Edge line color represents direction of the QTLs, green for higher levels in Sha; blue for higher levels in Bay-0. Line width indicate LOD scores. Detailed results comparing overlapping QTLs based on 95% confidence interval between models are shown in Supplementary table1.

Figure 3: Effect sizes for each individual developmental stages are plotted against the derived LOD score. A: normalized allelic effect size per environment against LOD scores from the genetic (G) component and B: normalized allelic effect size per environment against LOD scores from the genetic x environmental interaction (G:E) component. Colors indicate the developmental stages: red = primary dormant (PD); blue = after-ripened (AR); green = 6 hours imbibed (6H); orange = seeds at radicle protrusion (RP).
Figure 4: Normalized metabolite changes during 4 developmental stages (PD, AR, 6H and RP). Each panel represents a single metabolite and contains information about environmental variation (green line plot, average over all lines within a single developmental stage) and genetic variation (blue lines represent the metabolite levels for lines carrying the Bay-0 allele for the most significant QTL and red lines those for the Sha allele carrying lines). QTL profiles for metabolites with either genetic (G) or genetic x environmental (G:E) variation are indicated at the bottom of each panel by a heat bar representing the 5 chromosomes and a false-color scale is used to indicate the QTL significance. For G QTLs, positive values (light and dark blue) represent a larger effect on the metabolite content for the Bay-0 allele, and negative values (light and dark red) represent a larger effect on the metabolite content for the Sha allele. Interpretation of the color scale for G:E QTLs is less intuitive because strong negative environmental effects can result in inversion of the QTL LOD score (e.g. gluconic acid). The presented effectplot (left line plot) shows the true allele effect. Environmental (E) variation is expressed as LOD score in the lower left corner. Depending on the most significant variation either, genetic (G) or interaction (G:E) effects are also indicated with LOD scores in the lower left corner below or above the heat bar respectively. A: L-Lysine showing only Environmental (E) variation; B: Fumaric acid: showing Genetic (G) variation; C: Malic acid showing both Environmental and Genetic variation (G+E); D: Gluconic acid showing interaction between Environment and Genetic variation (G:E).

Figure 5: Number of significant QTLs plotted against the genetic location. Metabolic QTLs are represented by the solid (genetic component; G) and dashed (genetic x environmental component; G:E) lines. Germination related QTLs (Joosen et al., 2012) are shown by the dotted line.

Figure 6: QTL confirmation for maltose using the heterogeneous inbred family (HIF) approach. Two QTL regions (top chromosome IV and top chromosome V) were analyzed using after-ripened (AR) seeds of lines HIF312 and HIF214 (A). The QTL profile for maltose (B) shows two significant QTLs (dashed line indicates the LOD 4 significance threshold). The lower panel (C) shows the parental levels for maltose and the confirmation for both QTLs by the segregating HIF lines (either fixed for Bay-
0 or Sha alleles at the heterozygous interval). Significant differences (t-test p<0.05) are indicated with * in-between the two contrasting samples.
Table 1. Comparing different experimental designs.

<table>
<thead>
<tr>
<th>design</th>
<th>Full block design</th>
<th>Single block design</th>
<th>Random design</th>
<th>GGG design</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Best power for G</td>
<td>Same power for G</td>
<td>Limited power for G</td>
<td>Optimal power for G</td>
</tr>
<tr>
<td></td>
<td>Most expensive</td>
<td>Less expensive</td>
<td>Less expensive</td>
<td>Less expensive</td>
</tr>
<tr>
<td>QTL</td>
<td>96</td>
<td>93</td>
<td>78(75)</td>
<td>81(67)</td>
</tr>
<tr>
<td>Interacting QTL</td>
<td>30</td>
<td>0</td>
<td>17(5)</td>
<td>27(12)</td>
</tr>
</tbody>
</table>

Comparison of different experimental designs to study G and G x E effects on classic phenotypes in four different conditions. Each environmental condition is indicated with different gradient of grey in the blocks. In total there are N (=164) genetically different RILs and the data was analysed in 4 different ways. The last two rows compare the number of QTLs for main genotype effect (G) and interacting effect (G x E) detected using different design strategies, respectively. The numbers in brackets indicate the QTLs that share confidence intervals (1.5 drop-off) with the full block design.
A HIF312

msat4.39
msat4.43
msat4.35
nga8

Bay-0
Sha
Heterozygous

HIF214

nga249
nga151
msat5.14
nga139
msat512110

B Maltose

LOD score

I II III IV V

C

Log10 peak area

Bay-0 Sha HIF312 Bay-0 HIF312 Sha HIF214 Bay-0 HIF214 Sha