Genetic basis of seed metabolome variation in Arabidopsis

Elise A.R Serin[1], Leo A.J Willems[1], Harm Nijveen[1,2], Basten Snoek[3,4], Henk W.M Hilhorst[1] and Wilco Ligterink[1]

INTRODUCTION AND AIM

Changes in the environment during seed development affects seed performance in a genotype-dependant manner[]. We seek to identify the genes and molecular mechanisms that regulate seed quality acquisition. We grew recombinant inbred lines (RILs) of an Arabidopsis Bayreuth x Shahdara population in standard and three moderate stressful conditions from flowering until seed harvest. We used phenomics, metabolomics and transcriptomics in combination with classical quantitative trait loci (QTL) analysis to identify key genetic factors interacting with the environment (QTL x E).

Generalized genetical genomics (GGG)

A GGG design[2,3] was used to investigate mQTL x E in a cost-efficient manner. This approach consists in creating sub-populations of non-overlapping RILs from the four initial populations, while maximizing the total genetic variation (Fig. 1). **GC-TOF-MS** and **RNA-seq** were performed on the parental lines and the RILs of the sub-populations. The sub-populations were analysed separately and combined to identify stable and environment-specific mQTLs as well as mQTL x E interactions.

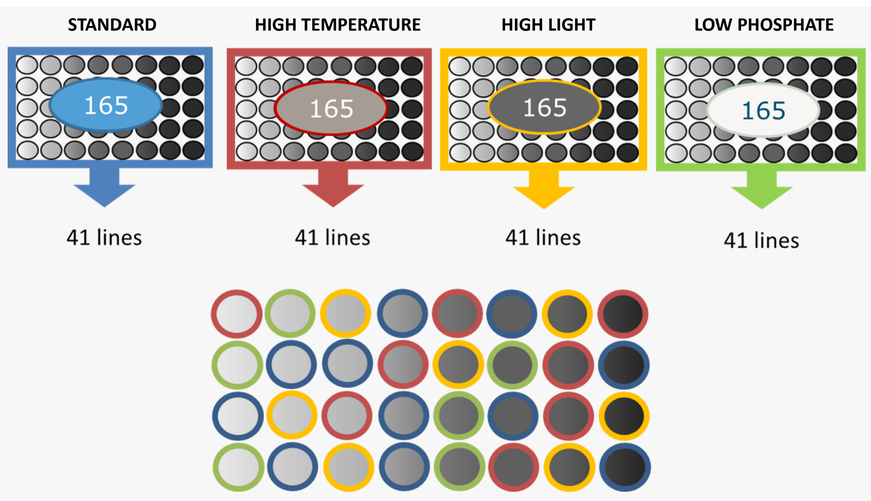

Figure 1. Schematic representation of the GGG approach. Grey scale color represents the total genetic variation.

RESULTS

G x E shapes dry seed metabolome

For the lines in the GGG design and the parental lines, dry seeds were used to extract and measure polar primary metabolites by GC-TOF-MS analysis. Overall 172 primary metabolites could be detected among which 71 could be unambiguously annotated. We were able to detect significant effects of the genotype, the environment and their interaction on the metabolite content of the parental lines (Bay-0 & Sha) grown under the different environments (Fig. 2). We also observed environment-specific metabolic profiles within the RILs (Fig. 3).

TCA cycle intermediates (fumarate, succinate) and GABA shunt metabolites (GABA, Alanine) did contribute to the separation of RILs developed at high temperature by the first component of the PCA (Fig. 3).

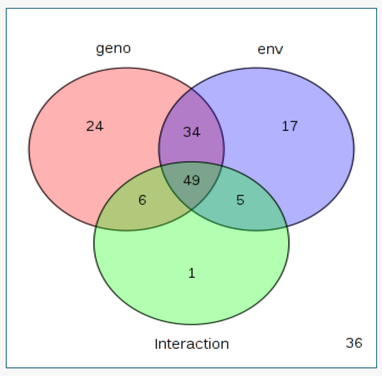

Figure 2. Two-way ANOVA reveals G, E and GxE effects on the metabolites detected in the parental lines.

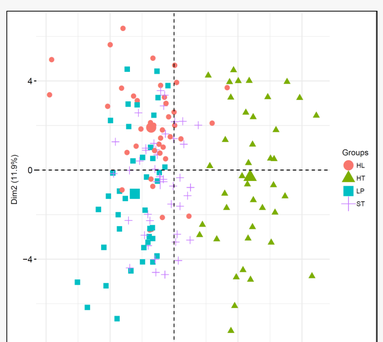

Figure 3. PCA analysis of the metabolites of RILs grown under the different environments.

Correlation network analyses

Metabolite correlation network analysis on the subpopulations of each environment revealed coordinated and environment-specific metabolic changes (Fig. 4).

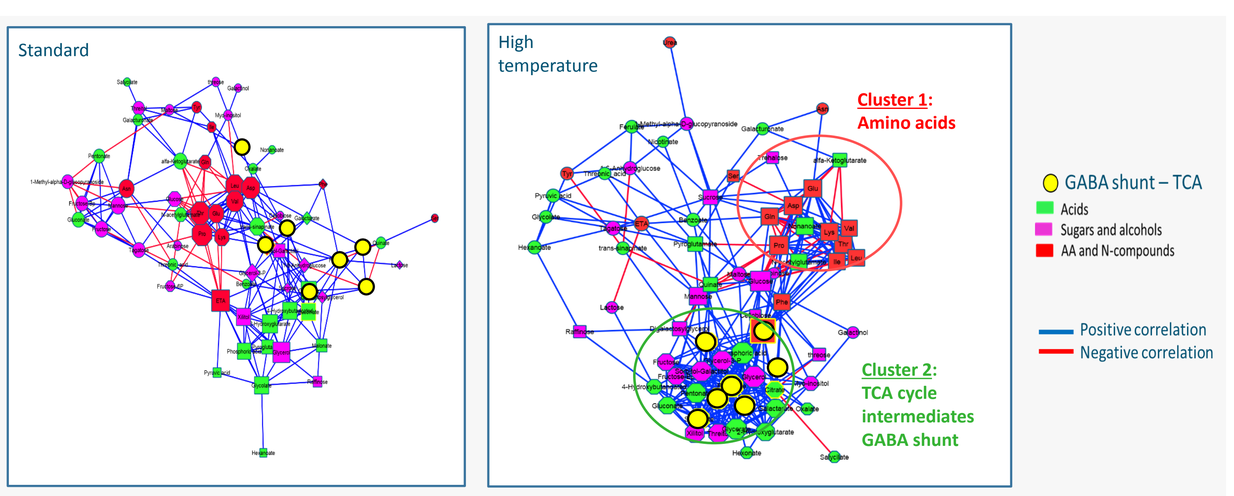

Figure 4: Condition dependent metabolite correlation networks. Only significant pearson correlations between metabolites are represented (FDR, q< 0.05) for the standard and high temperature network.

mQTL analyses

QTL analysis for the variation of the metabolite content across the combined RILs resulted in many mQTLs (Fig. 5). The mQTL hotspot on chromosome II (Fig. 5 a,b) contains mQTLs for metabolites found in cluster 2 of the HT network (Fig. 4). mQTLs for myo-inositol and galactinol were found in other sub-populations, while other mQTLs (e.g fumarate, succinate, glycine etc.) were found specifically in the HT sub-population (Fig. 5).

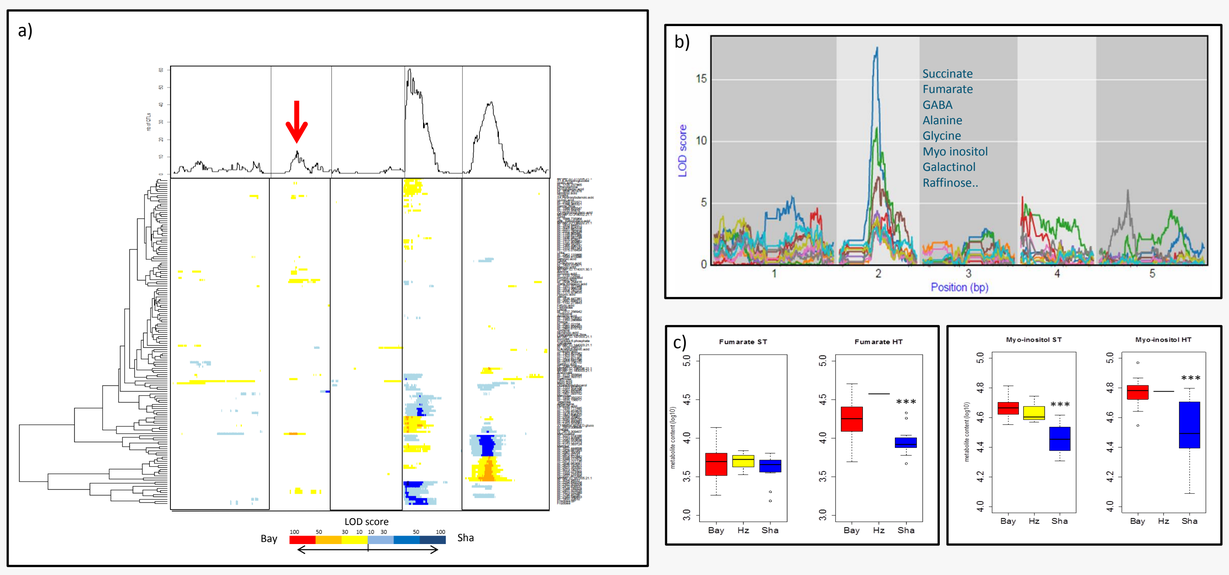

Figure 5: a) Heatmap of the mQTLs detected for the 172 metabolites. 3 major mQTL "hotspots" can be identified. The predominant hotspots are likely the known AOP and MAM loci on chromosome 4 and 5, respectively. b) metabolites with an mQTL in the hotspot region on chromosome II. c) Differences in the effect of the allele at the hotspot's nearest locus in the ST and HT sub-populations (p< 0.05).

Candidate genes

At2g22240: Inositol-3-phosphate synthase isozyme 2

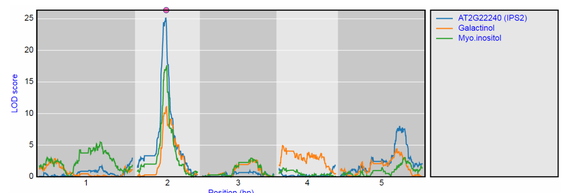

Figure 6: Co-locating mQTL and eQTL.

eQTL data obtained on the same RILs are used to identify putative regulatory genes.

As an example, mQTLs for galactinol and myo-inositol overlap with an eQTL for At2g22240 (Fig. 6). The known function of this gene in the Inositol phosphate pathway makes it the most likely candidate underlying these mQTLs.

CONCLUSIONS

- The GGG design efficiently detects QTLs for main genetic effects
- Correlation network analysis reveals clustered metabolic changes that can aid in QTL mapping analysis
- The combined analysis of mQTLs and eQTLs enables the identification of key regulators of seed quality

Laboratory of Plant Physiology
Droevendaalsesteeg 1, 6708PB Wageningen, The Netherlands
Contact: elise.serin@wur.nl
www.wageningenseedlab.nl

1 Wageningen Seed Lab, Laboratory of Plant Physiology, Wageningen University, The Netherlands
2 Laboratory of Bioinformatics, Wageningen University, The Netherlands
3 *C.elegans* group, Laboratory of Nematology, Wageningen University, The Netherlands
4 Theoretical Biology and Bioinformatics, Utrecht University, The Netherlands

References:
[1] He et al. 2014, **Interaction between parental environment and genotype affects plant and seed performance in *Arabidopsis***
[2] Li et al. 2008, **Generalizing genetical genomics: getting added value from environmental perturbation**
[3] Joosen et al. 2013, **Identifying Genotype-by-Environment Interactions in the Metabolism of Germinating Arabidopsis Seeds Using Generalized Genetical Genomics**