Genetic and maturation environmental modulation of tomato dry seed metabolites

Nafiseh Geshnizjani[1], Leo Willems[2], Henk Hilhorst[3], and Wilco Ligterink[3]

[1]Wageningen University and Research
[2]WUR
[3]Wageningen Universiteit

July 21, 2020

Abstract

Each seed phenotypic trait may be influenced by its metabolite composition which varies during seed development and -maturation. Therefore, the metabolic components of dry seeds may represent the maternal environment under which seeds developed and matured. The natural variation of seed metabolite composition has been previously studied and several QTLs were detected regulating this variation. Here a generalized genetical genomics (GGG) approach was used for the metabolic analysis of a recombinant inbred (RIL) population obtained from a cross between two tomato species: Solanum lycopersicum and Solanum pimpinellifolium. The RILs and parental lines were grown in two maternal environments: high phosphate and low nitrate nutrition. A correlation analysis of metabolite composition and seed phenotypic traits indicated several relations between metabolite contents and seed quality traits such as seed size, seed weight and seed germination percentage. Seed size and -weight exhibited a positive correlation with several amino acids and some intermediates of the TCA cycle, such as succinate, citrate and malate. By performing metabolic correlation analysis and also generating metabolite networks and combining these with QTL analysis, several relevant mQTLs were identified. We showed that dry seed metabolites and QTLs regulating them are modulated by genetic factors, maternal environment and their interaction.

Introduction

The maternal environment in which seeds develop and mature have a profound influence on seed properties such as germination vigour. The sink-source connection between the mother plant and the seeds allows the seeds to accumulate reserves required for seed germination and seedling growth (Baud *et al.* 2008). Metabolites such as amino acids, sugars and organic acids play a vital role in the different stages of seed development such as maturation, desiccation and germination (Borisjuk *et al.* 2004; Fait *et al.* 2006). During seed maturation, the content of these metabolites in seeds decreases and storage reserves, including starch, oil and seed storage proteins increase (Fait *et al.* 2006; Galili *et al.* 2014). It has also been shown that the subsequent metabolite content and composition of dry seeds may reflect the maturation environments in which they developed (He *et al.* 2016). For example, in different species it has been reported that nitrogen related metabolites such as asparagine, allantoin and GABA show a lower content in seeds developed under low nitrate maternal environments (Geshnizjani *et al.* 2019; He*et al.* 2016). Although many studies have been performed related to the effect of maternal environments on dry seed metabolic content, more information is required to understand the genetic and molecular mechanisms governing the metabolic changes in response to the maternal environment.

In general, each observed phenotype in plants is the consequence of different cellular processes such as gene transcription, protein translation and, finally, metabolite production (Kooke & Keurentjes 2011). Therefore, genetic variation is not only confined to phenotypic traits such as seed and seedling quality traits. Many

studies have revealed that metabolite composition and content, which play a very critical role in plant growth and development, is also controlled by genetic variation within a plant species (Windsor *et al.* 2005). The existing natural variation for both phenotypic traits and metabolite content is displayed by a continuous distribution, considered as quantitative variation. Such variations are often regulated by multiple loci and can be detected in mapping populations like recombinant inbred line (RIL) populations where the different loci are known as phenotypic or metabolite quantitative trait loci (QTLs and mQTLs, respectively) (Keurentjes & Sulpice 2009; Lisec *et al.* 2008). Many QTL analyses have been performed in seeds and many QTLs that regulate complex quantitative traits such as seed germination characteristics, seed size, seedling traits as well as seed metabolites have been described (Kazmi *et al.* 2012; Kazmi *et al.* 2017; Khan*et al.* 2012; Schauer *et al.* 2006).

Plants are a rich source of biochemical compounds that are mainly contributing to plant development, adaptation and final appearance and yield (Binder 2010). Therefore, the quantitative variation of these metabolites may have an influence on different physiological traits like seed germination and seedling establishment. The integrative analysis of metabolites and genetics has provided valuable information and knowledge on how natural variation regulates metabolite levels and their subsequent effect on growth of plants and their adaptation and how this knowledge can be used in plant breeding (Kliebenstein 2009).

Genetical genomics in which QTL analysis is integrated with proteomics, transcriptomics and metabolomics has provided in-depth understanding of molecular mechanisms regulating complex traits (Jansen & Nap 2001; Keurentjes *et al.* 2006; Kliebenstein *et al.* 2006; Schauer*et al.* 2006). Nonetheless, more advanced approaches are required for further determination of the complexity of quantitative traits. In addition to genotype (G), molecular networks are also influenced by the environment (E) and the interaction between genotype and the environment (G×E). Thus, the incorporation of different environments in genetic studies is a prerequisite for comprehensive perception of the regulation of molecular mechanisms. Li, Breitling and Jansen (2008) proposed a new strategy which is called generalized genetical genomics (GGG) by which both genetic and environmental perturbations can be studied. This approach allows QTL analysis governing the interesting molecular traits under consideration of multiple environments. It is a cost-effective method to not only determine the genotype but also the environmental effects and their interaction for detected QTLs (Li, Breitling & Jansen 2008). In principle, by creating similar subpopulations of RILs and subjecting each of these to a different environment, G, E and G×E effects can be investigated in a cost-effective experimental design (Joosen *et al.* 2013).

Although the QTLs governing dry seed metabolite content have been previously detected in many plants, including tomato (Kazmi *et al.* 2017; Toubiana *et al.* 2012), the effect of G×E interactions has been studied to a much lesser extent (Albert *et al.* 2016; Kazmi *et al.* 2017; Rosental *et al.* 2016). In this study we used a RIL population derived from a cross between two tomato species:*Solanum lycopersicum* (cv. Moneymaker) and *Solanum pimpinellifolium* (Voorrips *et al.* 2000). We have exploited the existing natural variation in this population to investigate how QTLs are influenced by the environment to which the mother plants are exposed. Moreover, metabolic profiling of the seeds which have matured in different environments will be useful to illustrate important metabolic differences that regulate the development and adaptation of plants (Joosen *et al.* 2013). By using a GGG approach we performed metabolite analysis for the RIL population and their parental lines, grown in high phosphate and low nitrate environments. By generating metabolite correlation networks and performing mQTL analysis, genetic and molecular aspects of seed metabolic changes in response to the maternal environments have been discovered.

Materials and methods

Maturation conditions and seed collection

One hundred lines of an F_7 RIL population obtained from*S. lycopersicum* cv. Moneymaker (MM) × *S. pimpinellifolium*(PI) accession G1.1554 (Voorrips *et al.* 2000) have been genotyped with 865 single nucleotide polymorphism (SNP) markers. The F_8 population was grown in two different nutritional maturation environments as previously described (Geshnizjani *et al.* 2020). The fully ripened fruits were collected and

the seeds were extracted and dried as previously reported (Chapter 2). Finally, the dry seeds were stored in paper bags at 13°C and 30% RH.

Generalized genetical genomics design (GGG)

The population of 100 tomato lines was divided into two sub-populations based on the distribution of parental alleles. By using the R-procedure DesignGG (Joosen *et al.* 2013; Li *et al.* 2009) the tomato lines were allocated to the suitable sub-population in a way that alleles show a similar distribution in both sub-populations as compared to the whole population (Kazmi *et al.* 2017).

Extraction and analysis of dry seed metabolites

The dry seed metabolites were extracted using the method as previously described by Roessner et al., (2000) with small changes. In short, 10 mg seeds of each tomato line was homogenized using a micro dismembrator (Sartorius) in a precooled 2 ml Eppendorf tube with 2 iron balls (2.5 mm). A solution of 700 µl methanol/chloroform (4:3) together with a standard (0.2 mg/ml ribitol) was added to each Eppendorf tube and mixed thoroughly. After 10 minutes of sonication 200 µl Milli-Q water was added to the samples followed by vortexing and centrifugation (5 min, 13,500 rpm). Then, the methanol phase was collected and transferred to a new 2 ml tube and the remaining organic phase was extracted again with 500 µl methanol/chloroform. The solution was kept on ice for 10 minutes and afterwards 200 µl Milli-Q water was added. Again after vortexing and centrifugation (5 min, 13,500 rpm), the methanol phase was collected and combined with the former collected phase and mixed well. A solution of 100 µl of this mix was transferred to a glass vial and dried overnight using a speedvac centrifuge at 35°C (Savant SPD1211).

The gas chromatography-time of flight-mass spectrometry (GC-TOF-MS) method which was previously described by Carreno-Quintero et al., (2012) was used for analysis of the dry seed metabolites. Detector voltage was set at 1600 V. Analysis of the raw data was performed using chromaTOF software 2.0 (Leco instruments). Furthermore, the Metalign software was used for further analysis such as aligning the mass signals (Lommen 2009). The peak threshold for noise was set to 2 and the output was loaded in Metalign Output Transformer (METOT; Plant Research International, Wageningen) and MSClust (Tikunov *et al.* 2012) was used to construct Centrotypes. The Centrotypes were identified by matching the mass spectra to an in-house-constructed library, to the GOLM metabolome database (http://gmd.mpimp-golm.mpg.de/) and to the NIST05 library (National Institute of Standards and Technology, Gaithersburg, MD, USA; http://www.nist.gov/srd/mslist.htm). The identification was based on spectral similarities and comparing the retention indices calculated by a third order polynomial function (Strehmel *et al.* 2008).

Statistical analysis (coefficient of variation, PCA and ANOVA analysis)

Within a population the absolute variation or dispersion per trait is defined as the standard deviation (σ). The relative variation called the coefficient variation (CV) for individual traits is the ratio of the standard variation to the mean (μ) of the lines in the population (CV = (σ/μ)*100). In this study we calculated CV for each metabolite in two nutritional maternal environments separately. The metabolomics data were log10 transformed and then used for further analysis such as ANOVA, principal component and correlation analysis. In order to indicate the effect of the genotype, environment and their interaction, ANOVA analysis was performed on the metabolite content of the parental lines, MM and PI, grown in different nutritional environments. The significance threshold of the p -value was set to 0.05. Principal component analysis (PCA) was performed on metabolic values of the RILs and the parental lines using the R-package "pcaplots".

Log2 ratio of metabolites between HP and LN (HP:LN) and metabolite profiles in both maternal environments (HP and LN) were investigated to identify the metabolic differences between two nutritional maternal environments with Metaboanalyst 3.0 (http://www.metaboanalyst.ca/faces/home.xhtml).

Correlation analysis and network construction

R-packages "MASS", "Hmisc", "VGAM", "gplots" and "graphics" (https://www.r-project.org/) were used for analysis and construction of the Spearman correlation between all known metabolites and also between

the metabolites and the seed phenotypic traits. In addition, by using the "rcorr" R-package for each sub-population the Spearman correlation between the known metabolites was analysed and the significance level of correlations was described as false discovery rate (FDR). Correlation values with FDR[?]0.05 were used to create a correlation network for each maternal environment by using Cytoscape v.3.4.0. The Network-Analyser tool of Cytoscape was used to obtain additional characteristics of the metabolic networks.

mQTL analysis

Log10 transformed data together with the tomato linkage map containing 865 SNP markers were used for performing QTL analysis using Rqtl v3.3.1 (Arends *et al.* 2010; Broman *et al.* 2003). We performed QTL analysis using the approach as previously described by Joosen et al., (2013) for Arabidopsis and Kazmi et al., (2017) for tomato, with small modifications. A model (Y = G + E + G×E +) was used for the whole RIL population to identify the effect of genotype (G), maternal environment (E) and their interaction (G×E). Furthermore, in order to identify the metabolic variation explained by the genetic component we carried out the QTL analysis for known metabolites in each sub-population with simple interval mapping (SIM) using MapQTL® 6.0 (Van Ooijen 2004). In both QTL analysis (Rqtl and MapQTL) 1000 permutation tests were applied to our data to estimate the LOD threshold at a significance level of 0.05.

Results and discussion

Principal component analysis

In this study we used 100 lines of a tomato recombinant inbred line (RIL) population derived from a cross between *Solanum lycopersicum* (cv. Moneymaker) (MM), and *Solanum pimpinellifolium*accession G1.1554 (PI) (Voorrips *et al.* 2000). The RILs, together with the parental lines, were grown in two different nutritional environments; high phosphate and low nitrate. The harvested seeds were used to measure the metabolite content in the dry mature seeds. The metabolites were measured by gas chromatography-time of flight-mass spectrometry (GC-TOF-MS) analysis in a specific GGG design. In total 118 primary metabolites were detected from which 58 could be identified. These identified metabolites were classified as amino acids, organic acids, sugars, sugar alcohols and some other compounds (**Table S1**).

Principal component analysis of the primary metabolites in the parental lines (MM and PI) indicated clear genetic effects as the two genotypes displayed different accumulation of metabolites in each maternal environment. In addition, maternal environmental effects were also observed within each genotype as metabolites accumulated differently in HP and LN (the two different maternal environments) (**Figure 1**).

Furthermore, analysis of significant changes in dry seed metabolites of the two species (MM and PI) between the different environments revealed that the metabolite contents were influenced to a higher extent in MM as compared to PI. In MM the level of 36 metabolites, mostly including amino acids and organic acids, had significantly changed between the environments, while for PI this number decreased to 26 from which 19 were common between the two species (**Table S2**). In general, PI as a wild tomato species is more tolerant to suboptimal environments (Kumar 2006; Rao *et al.* 2013; Rodríguez-López *et al.* 2011) which might be the reason why it does not need to modify its metabolites to a high extent to cope with a changing environment. While in the case of MM, which is considered as a domestic sensitive accession, many metabolites may have to be altered in order to allow it to deal with environmental stresses.

Metabolites are not only influenced by the genetic background of the seeds but also by the environment under which seeds develop and mature and the interaction between genotype and environment (G×E) (He *et al.* 2016). In our study, a wide range of the metabolites in the parental lines were also significantly influenced by the genotype, nutritional environment and their interaction (**Table S3**). In total 15% of the annotated metabolites did not significantly change in our analysis and 57, 60 and 53% of the metabolites were influenced by genotype, environment and G×E, respectively.

mQTL analysis

The substantial variation caused by G, E and G×E may be an indication of a sophisticated regulation

of metabolites in developing tomato seeds. Thus, metabolite content of the seeds must be considered as a complex trait which is likely regulated by multiple quantitative trait loci (QTLs) (Keurentjes & Sulpice 2009; Lisec *et al.* 2008).

To detect these mQTLs, we analysed metabolite profiles in the dry mature seeds of all the RILs. PCA analysis of this data revealed a clear separation between the known metabolites of the seeds grown in the two nutritional maternal environments (**Figure 2A**). Twenty-three and 15% of the total variation was explained by PC1 and PC2 respectively. Similar results were obtained after analysis of all 118 detected metabolites (**Figure S1A**). Both PCA plots show that predominantly PC2 explained the variation related to the maternal environment. The loading plot of the PCA showed that the main components contributing to PC2 were organic acids including glycolate, glycerate, malonate and succinate and amino acids such as glutamate, serine, threonine and asparagine (**Figure 2B, Table S4**).

Metabolite profile and correlation in HP and LN

ANOVA analysis of the metabolites in the two maternal environments indicated that contents of 36 of 58 known metabolites were significantly changed between HP and LN maternal conditions. These metabolites included mostly amino acids such as serine, pyroglutamate and GABA and organic acids including TCA cycle intermediates such as galacterate, malate, succinate and malonate (**Table 1**).

In general Log2 ratios of HP:LN in the RIL population showed that most of the metabolites had a higher level in seeds from the HP environment as compared to the LN environment (**Figure 3, Table 1**).

The metabolite profiles of seeds from the two maternal environments showed that seeds grown in HP conditions contained a higher amount of metabolites, such as many sugars, amino acids and organic acids (**Figure 3, 4**). A high metabolite level was predominantly observed for GABA, sugars (maltose, fructose and glucose), organic acids including benzoate, salicylate, glycerate and also some TCA cycle intermediates such as galacterate, malate, malonate and succinate. The TCA cycle, including several catabolic reactions, plays a very critical role in energy metabolism in plants. In addition, it contributes in many other ways to the metabolome by being involved in a large number of metabolic networks (Araujo *et al.* 2012). In plants, it has been shown that amino acids such as leucine, iso-leucine and valine can be degraded into new products which can be used as precursors for the TCA cycle to provide additional energy for plants. The rate of the degradation of these compounds can be increased due to sub-optimal conditions such as abiotic stresses (Binder 2010).

Both HP and LN maternal environments could be sub-optimal environments for plant growth and seed development. Our results revealed that most of the amino acids were not significantly altered between the two maternal environments; however some of them such as GABA, pyroglutamate, glycine, leucine and aspartate showed significantly lower values in seeds grown in LN. This result was consistent with findings for tomato and Arabidopsis plants that reported a general decrease of amino acid levels under low nitrate conditions (Tschoep *et al.* 2009; Urbanczyk-Wochniak & Fernie 2004). GABA is one of the amino acids which frequently shows higher levels under stress conditions (Michaeli*et al.* 2011; Renault *et al.* 2011; Shelp, Bown & McLean 1999). GABA is an amino compound which is produced via the so-called GABA shunt pathway which has a primary role in keeping a balance in central C/N metabolism (Bouche & Fromm 2004). It has been shown that GABA levels increased rapidly under stressed growth conditions. Thus, GABA is thought to be involved in the tolerance of plants to sub-optimal environments (Fait *et al.* 2008; Kinnersley & Turano 2000; Renault *et al.* 2011). In our study we have observed that the GABA content was lower in seeds developed in LN conditions in comparison with those of the HP maternal condition. Our findings confirmed previous studies, where seeds also showed low amounts of GABA under LN maternal conditions (He *et al.* 2016); (Geshnizjani *et al.* 2019).

It has previously been reported that nitrate starvation resulted in a decrease in TCA cycle intermediates in tomato (Urbanczyk-Wochniak & Fernie 2004) and Arabidopsis (Tschoep *et al.* 2009) leaves. We also found that in comparison to the HP condition, seeds developed under LN possessed a lower amount of TCA cycle intermediates such as malate, succinate and malonate (**Figure 4**). This could be an indication of higher

consumption of TCA cycle intermediates to produce more energy under LN to allow the plants to survive and continue growth.

Metabolite correlation networks

In general, correlations between metabolites can be used to assist in unravelling the biological basis of variation caused either by different environments or genetic backgrounds (Ursem *et al.* 2008). In order to understand the correlation between metabolite contents within the RIL sub-populations and how their interaction is influenced by the nutritional maternal environment, pairwise Spearman correlation analysis was performed between the metabolites. For each environmental condition, correlation analysis of all 118 detected metabolites has been performed and a correlation heatmap was generated (**Figure S2, Table S5**). The results showed that most of the unknown metabolites are highly correlated with annotated metabolites such as amino acids and organic acids including TCA cycle intermediates. Only known metabolites that showed significant correlations (FDR[?]0.05) were selected for constructing correlation networks (**Figure 5, Table 2**). By using the network approach, the correlation between metabolites within each sub-population as a result of similar genetic regulation can be visualised, while different metabolic patterns in between the different maternal environments could provide more insight into the influence of environment and GxE on regulation of metabolites. Correlation networks have often been used in metabolomics studies (Morgenthal, Weckwerth & Steuer 2006; Steuer*et al.* 2003) to provide additional information to multivariate approaches which have been described previously (Graffelman & van Eeuwijk 2005). In our study, the correlation network for the HP maternal environment contains in total 395 significant correlations (edges) between 56 metabolites (nodes). The HP condition resulted in a network with higher density (0.256) as compared to LN, which had in total 238 edges and 51 nodes (**Table 2**). In general, the network related to the HP environment showed higher levels of some attributes such as range of node degree, number of nodes and edges, network density and average number of neighbours by possessing more metabolite connections and correlations (**Table 2**). This higher connectivity in the network could be related to the overall higher metabolic levels under this specific condition. In our study dry seed metabolites were connected more under the HP condition, in comparison with LN, which indicates that the regulatory mechanisms under HP conditions induce several changes in metabolism. These metabolic changes could assist plants to cope with sub-optimal growing conditions and may result in acclimation of the plant (Hochberg *et al.* 2013).

The most highly connected metabolites in each condition can be found in**Table S6** . Under LN, mainly amino acids are highly correlated with each other and thus could be predominantly involved in metabolic changes due to LN conditions (**Figure 5A**). However, under HP maternal condition, in addition to the amino acids such as alanine, glycine, serine and threonine, some of the TCA cycle intermediates including malate, fumarate and succinate are also highly connected (**Figure 5B**). In both environments we observed strong correlation between metabolites within the same category such as amino acids. Such a consistent correlation observed in both environments suggested that these metabolites are mainly under genetic control and not much influenced by the environment or GxE interactions. In our results under HP conditions glycine showed a strong correlation with malate (one of the TCA cycle intermediates, R = 0.6, FDR = 0.00021) while we could not find it back in the LN network. Such different network topologies indicate a strong environmental effect on the correlation between these metabolites. These examples show that the correlation networks and the differences amongst them may provide imperative information to understand the molecular basis of metabolic changes (Schauer *et al.* 2006).

Correlation of metabolites within the whole RIL population

Spearman correlation analysis was performed and a correlation matrix was generated between all pairs of known metabolites across the whole RIL population. The results revealed that some metabolites are highly correlated with each other (**Figure 6**). Except a few exceptions, all amino acids cluster together. They showed a high degree of correlation of mostly greater than 0.5 with p -values of less than 0.001 (**Table S7**). Such a high positive significant correlation could be an indication of a preserved metabolism of amino acids in seeds. It has previously been reported that the metabolism of amino acids in seeds might be regulated by post-transcriptional regulators in order to regulate the distribution of nitrogen (Kazmi *et al.*2017; Toubiana

et al. 2012). We also observed a high number of significant correlations between amino acids and TCA cycle intermediates such as citrate, malate, fumarate and succinate. Such a correlation between amino acids as a nitrogen (N) source and TCA intermediates as carbon (C) metabolites, indicates a maintained crosstalk between N and C metabolism in the seeds (**Figure 6, Table S7**). A similar crosstalk has been previously suggested for different species including Arabidopsis and tomato (Gutierrez *et al.* 2007; Kazmi *et al.* 2017; Nunes-Nesi, Fernie & Stitt 2010; Stitt & Fernie 2003).

We have also detected significant correlation between galactinol and myo-inositol (R=0.53 and *p* -value=1.06E-6) (**Figure 6, Table S7**). These metabolites are classified as sugar alcohols which have been reported to be involved in responses of seeds to stressful environments, such as low temperature (He *et al.* 2016).

Correlation between metabolites and seed phenotypic traits

In order to assess the relationship between metabolites and seed phenotypic traits, seed performance phenotypes which were previously assessed for the same seeds (Geshnizjani *et al.* 2020), were integrated into the metabolic correlation matrix (**Figure 7**). We found many positive and negative correlations between metabolites and phenotypic traits. The results revealed that seed size and weight are positively correlated with most of the amino acids and TCA cycle intermediates such as succinate, citrate and malate (**Figure 7, Table S8**). The strongest positive correlation was found between seed size and amino acids including pyroglutamate, leucine and isoleucine (R[?]0.4, *p* -value<0.0001). Among all the seed germination traits maximum germination percentage (G_{max}) showed the highest number of significant correlations with metabolites of which most are negative. G_{max} under osmotic stress (mannitol and NaCl) has a significant positive correlation with 2-ketoglutarate which is one of the TCA cycle intermediates, involved in supplying the required energy for seed germination (**Table S8**). G_{max} under optimal and sub-optimal germination environments showed strong negative correlation with many of the amino acids (such as pyroglutamate, GABA, methionine and leucine), organic acids (glycerate and malonate) and TCA cycle intermediates (malate and succinate).

Amino acids, are the precursors of protein synthesis and also precursors of some TCA cycle intermediates (e.g. citrate and succinate), serve as energy generation units for embryo growth as well as radicle protrusion (Lehmann & Ratajczak 2008; Ratajczak *et al.* 1996; Rosental, Nonogaki & Fait 2014). Since energy and proteins are two elements supporting germination, such a negative correlation between them and germination of tomato seeds is not expected. However, our results are in accordance with several foregoing studies which reported that accumulation of amino acids, such as methionine, lysine and GABA, may cause inhibition of seed germination (Amir 2010; Angelovici *et al.* 2011). In some other reports, amino acids were considered as one of the biological methods to control weeds since the external application of many amino acids decreased the seed germination percentage for some species such as broomrape (Vurro *et al.* 2006; Wilson & Bell 1978). Such a negative effect of amino acids on seed germination could be related to accumulation of certain amino acids in the seeds and subsequent reduction of some other metabolites such as TCA cycle intermediates which may play vital roles in seed germination (Angelovici*et al.* 2011; Rosental *et al.* 2016). For example, the biosynthetic pathway of lysine uses pyruvate which is the central component of the TCA cycle. Depletion of pyruvate from the TCA cycle will ultimately result in a decrease in the production of TCA cycle intermediates. Hence, such a decrease in TCA cycle input results in declined levels of available energy, which in turn negatively affects seed germination (Angelovici *et al.* 2011; Day *et al.* 1994; Shedlarski & Gilvarg 1970). A strong negative correlation was found between G_{max} in water and methionine content of the seeds (R=0.42, *p* -value<0.001). Similar results have been found in different species such as lettuce (Wilson & Bell 1978) and tomato (Rosental *et al.* 2016). Feedback inhibition of increased methionine on the upstream enzymes activity such as cystathionine γ-synthase (CGS) has been reported before (Chiba *et al.* 2003; Rosental *et al.* 2016). Hence high methionine content of seeds may limit the synthesis of sulfur-rich proteins which subsequently results in the reduction of seed germination (Amir 2010). However, our findings seem in contrast with a few other studies in which a high level of methionine did not lead to a decrease of germination which indicated that methionine was not negatively correlated with germination (Amir, Han & Ma 2012; Gallardo *et al.* 2002).

We also performed a correlation analysis between metabolites and seed phenotypic traits within each tomato RIL sub-population and two correlation heatmaps were generated (**Figure S3**). In general, substantial differences were not observed between the two maternal environments (HP and LN); however, correlations appeared stronger within HP as compared to LN conditions and some correlations were specifically observed in one of the environments. For example, the positive correlation observed between many amino acids and seed size and weight were either lost at LN or were not as strong as what was observed at HP (**Figure S3**). In addition, a limited number of metabolites (e.g. galactarate) showed a significant positive correlation with most of the phenotypic traits in LN; however, the same metabolite showed a weak negative correlation with the same seed phenotypic traits in HP (**Figure S3**). An association of germination percentage and metabolic content of the dry seeds may raise the possibility to predict germination behaviour using the metabolic signature of the dry seeds (Rosental, Nonogaki & Fait 2014).

mQTL profiling of the tomato RIL population

The calculation of the coefficient of variation (CV) showed that most of the metabolites possess a CV value higher than 40%, which indicates that there is considerable variation within the RIL population for the metabolite levels in the dry seeds (**Figure 8, Table S9**). In order to investigate if such a high level of variability within metabolites could be explained by differences in alleles and genetic factors, a metabolic quantitative trait locus (mQTL) analysis was performed with the obtained metabolite data. Each metabolite is in general controlled by several pathways and regulators. Thus, as expected, we hardly identified metabolites for which a single genetic locus significantly explained the metabolite levels.

In our study we performed mQTL analysis for each maternal environment to evaluate the genetic variation within each sub-population. Furthermore we used the whole set of RILs to detect mQTLs explained by a genetic component (G) and the genotype by environment interaction (G×E).

We identified mQTLs across all conditions (**Table 3**). Regarding the sub-populations 66 and 129 mQTLs were detected for seeds from LN and HP environments, respectively. The heatmap of the LOD profiles and characteristics of the mQTLs in each environment are presented in**Figure S4** and **Table S10** , respectively. In both maternal environments several mQTLs were detected which were hardly detected for other metabolites. For example, in the seeds developed under HP conditions a single strong QTL on chromosome 9 was detected, regulating asparagine. Another independent significant QTL was identified on the top of chromosome 11 for phenylalanine under the same environmental conditions (**Figure S4A, Table S10**). Detection of such specific mQTLs in our data reveals the tight and independent genetic regulation of metabolite biosynthesis in seeds (Keurentjes*et al.* 2008). Under the same maternal condition some organic acids such as benzoate, gluconate, glycerate and glycolate mapped to a similar position on chromosome 5 (**Figure S4A, Table S10**). On chromosome 9, we detected mQTLs for TCA cycle intermediates including citrate and malate which were co-locating with the one regulating F6P as one of the precursors of the TCA cycle. There is also a QTL on the top of chromosome 1 affecting amino acids in seeds from the HP environment. Despite the strong correlation that has been found between amino acids and TCA cycle intermediates in seeds from HP conditions, no co-located QTLs were identified for them. This might be due to several smaller QTLs regulating variation of the metabolites, each of them explaining a small part of the variation and therefore not reaching the threshold LOD score. Regarding the seeds grown in the LN maternal environment we found more than one QTL for some of the metabolites such as GABA, citrate and malate. The vital role of these metabolites has been reported in relation with the alleviation of environmental stress effects (Kaplan*et al.* 2004; Kinnersley & Turano 2000; Krasensky & Jonak 2012; Obata & Fernie 2012). For the LN environment many of the amino acids have co-locating QTLs at the bottom of chromosome 4 and in the middle of chromosome 5 (**Figure S4B**). Such strong co-locating QTLs for amino acids was expected since they showed a high connection in the correlation network of the LN environment (**Figure 5A**). In general, such co-localizing QTLs for metabolites suggest that, in addition to the single independent QTLs regulating metabolite contents, some general regulatory loci and genes are involved in the regulation of metabolite synthesis (Keurentjes *et al.* 2008).

Combining the sub-populations and using the whole set of RILs leads to an increase in the number of detected

QTLs with 382 and 146 QTLs for G and G×E effects, respectively. An overview of the detected QTLs is provided by the heatmap of the LOD profiles (**Figure 9**). On the top and bottom of chromosome 4 there are two QTLs that explain the variation for many amino acids such as aspartate, GABA, glutamine, methionine, serine and threonine. Similarly, a co-located QTL was detected for galactarate and malate on chromosome 10 (**Figure 9**). Co-localization of these mQTLs is not surprising since galactarate is the precursor of 2-oxoglutarate and 2-oxoglutarate is one of the intermediates of the TCA cycle and is generally converted to malate in a couple of subsequent reactions. Our results show that myo-inositol and galactinol are highly associated with each other and closely grouped together. Therefore, it is not surprising that they both have a co-locating QTL on chromosome two (**Figure 6, Figure 9**). The robust correlation between raffinose pathway metabolites including galactinol and myo-inositol has also been reported for seeds of other species that developed under environmental stress (Cook *et al.*2004; He *et al.* 2016). These metabolites are known for their protective role for cellular structures of embryos during seed development and desiccation (Taji *et al.* 2002). Furthermore, they are able to play a key role in protecting plants from the effects of stress resulting from reactive oxygen species (ElSayed, Rafudeen & Golldack 2014). Some of the organic acids including gluconate, glycerate and glycolate, together with two of the TCA cycle intermediates (malate and succinate), had a co-locating QTL on chromosome 9 (**Figure 9**). Glutamate and GABA showed a shared QTL on chromosome 4 which has previously been detected in the same population developed under standard conditions (Kazmi *et al.* 2017). Metabolites belonging to the same functional class are often highly correlated and can have co-locating mQTLs (Kazmi *et al.* 2017). Although several mQTLs were detected at similar positions, in general more co-located mQTLs would be expected due to the strong correlation that has been observed between the metabolites. This could be related to the fact that several small QTLs are involved in regulation of the metabolites and each of them is explaining only a small part of their variation. Such small QTLs are likely to escape the QTL significant threshold in the QTL analysis (Keurentjes *et al.* 2008).

A few mQTLs co-located with the phenotypic QTLs that have been detected in a previous study (Geshnizjani *et al.* 2020). For instance, the QTLs on the middle of chromosome 10 affecting galactarate and malate co-located with ones influencing uniformity of germination (U_{8416}) at different germination conditions, such as high temperature, mannitol, water and NaCl. In addition, the QTL on chromosome 9, which is specifically regulating methionine, is located at the same position as QTLs affecting seed size, seed weight and fresh and dry weight of the seedlings. Despite the many strong correlations between metabolites and phenotypic traits (**Figure 9**), we could hardly detect co-locating QTLs for them. This might be due to the fact that each of the phenotypes may not be correlated with a specific metabolite but with a group of metabolites and thus the final metabolic balance between the groups of metabolites could affect phenotypic traits such as G_{max}.

Conclusion

In this study we performed GC-TOF-MS metabolite profiling of a tomato RIL population and their parental lines grown in high phosphate and low nitrate environments. Our results show clear genetic variation at the metabolite level between the two parental lines, where the maternal nutritional environment was also introducing variation within each genotype. Elucidation of genetic and molecular aspects of metabolic changes of seeds as a response to different maternal environments was carried out by using metabolite correlation networks, followed by mQTL analysis. In general the HP environment induced more metabolic changes as compared to the LN environment. Correlation of metabolites within the whole RIL population revealed a crosstalk between N and C metabolism in which significant correlations were observed between amino acids and TCA cycle intermediates. Besides mQTLs detected in the individual environments and the genetic effects, many mQTLs were detected for G×E. In spite of the strong correlations found between metabolites and phenotypic traits, the detected mQTLs were hardly co-located with the ones affecting phenotypic traits. This might be caused by the fact that not a single metabolite, but a group of metabolites together influence the phenotype. This study has provided novel insights towards better understanding of the effect of maternal environment on tomato seed and seedling performance by combining various physiological, omics and genetical analyses. In addition to the new insights that have been provided in this study, more in-depth investigations are needed to further elucidate the regulation of the dry seed metabolome under different nutritional environments and its influence on seed and seedling performance.

Acknowledgements

This work was supported by Applied and Engineering Sciences (TTW), a division of the Netherlands Organization for Scientific Research (NWO) (LW, WL).

Conflict of interest

The authors declare no conflicts of interest.

Author contributions

HWMH and WL conceived the study. NG and LAJW performed the experiments. LAJW and NG analysed the data. NG and WL wrote the paper with help from all co-authors.

References

Albert, E., Gricourt, J., Bertin, N., Bonnefoi, J., Pateyron, S., Tamby, J.-P., Bitton, F. & Causse, M. (2016) Genotype by watering regime interaction in cultivated tomato: lessons from linkage mapping and gene expression. *Theoretical and applied genetics* **129,**395-418.

Amir, R. (2010) Current understanding of the factors regulating methionine content in vegetative tissues of higher plants. *Amino acids* **39,** 917-931.

Amir, R., Han, T. & Ma, F. (2012) Bioengineering approaches to improve the nutritional values of seeds by increasing their methionine content.*Molecular breeding* **29,** 915-924.

Angelovici, R., Fait, A., Fernie, A. R. & Galili, G. (2011) A seed high-lysine trait is negatively associated with the TCA cycle and slows down Arabidopsis seed germination. *New Phytologist* **189,**148-159.

Araujo, W. L., Nunes-Nesi, A., Nikoloski, Z., Sweetlove, L. J. & Fernie, A. R. (2012) Metabolic control and regulation of the tricarboxylic acid cycle in photosynthetic and heterotrophic plant tissues. *Plant, Cell & Environment* **35,** 1-21.

Arends, D., Prins, P., Jansen, R. C. & Broman, K. W. (2010) R/qtl: high-throughput multiple QTL mapping. *Bioinformatics* **26,**2990-2992.

Baud, S., Dubreucq, B., Miquel, M., Rochat, C. & Lepiniec, L. (2008) Storage reserve accumulation in Arabidopsis: metabolic and developmental control of seed filling. *The Arabidopsis Book* e0113.

Binder, S. (2010) Branched-chain amino acid metabolism in Arabidopsis thaliana. *The Arabidopsis Book* e0137.

Borisjuk, L., Rolletschek, H., Radchuk, R., Weschke, W., Wobus, U. & Weber, H. (2004) Seed development and differentiation: a role for metabolic regulation. *Plant Biology* **6,** 375-386.

Bouche, N. & Fromm, H. (2004) GABA in plants: just a metabolite? *Trends Plant Sci* **9,** 110-115.

Broman, K. W., Wu, H., Sen, Ś. & Churchill, G. A. (2003) R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19,**889-890.

Chiba, Y., Sakurai, R., Yoshino, M., Ominato, K., Ishikawa, M., Onouchi, H. & Naito, S. (2003) S-adenosyl-L-methionine is an effector in the posttranscriptional autoregulation of the cystathionine γ-synthase gene in Arabidopsis. *Proceedings of the National Academy of Sciences***100,** 10225-10230.

Cook, D., Fowler, S., Fiehn, O. & Thomashow, M. F. (2004) A prominent role for the CBF cold response pathway in configuring the low-temperature metabolome of Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America***101,** 15243-15248.

Day, D. A., Millar, A. H., Wiskich, J. T. & Whelan, J. (1994) Regulation of alternative oxidase activity by pyruvate in soybean mitochondria. *Plant physiology* **106,** 1421-1427.

El Sayed, A., Rafudeen, M. & Golldack, D. (2014) Physiological aspects of raffinose family oligosaccharides in plants: protection against abiotic stress. *Plant Biology* **16,** 1-8.

Fait, A., Angelovici, R., Less, H., Ohad, I., Urbanczyk-Wochniak, E., Fernie, A. R. & Galili, G. (2006) Arabidopsis seed development and germination is associated with temporally distinct metabolic switches. *Plant physiology* **142,** 839-854.

Fait, A., Fromm, H., Walter, D., Galili, G. & Fernie, A. R. (2008) Highway or byway: the metabolic role of the GABA shunt in plants. *Trends in plant science* **13,** 14-19.

Galili, G., Avin-Wittenberg, T., Angelovici, R. & Fernie, A. R. (2014) The role of photosynthesis and amino acid metabolism in the energy status during seed development. *Frontiers in plant science* **5,** 447.

Gallardo, K., Job, C., Groot, S. P., Puype, M., Demol, H., Vandekerckhove, J. & Job, D. (2002) Importance of methionine biosynthesis for Arabidopsis seed germination and seedling growth. *Physiologia plantarum* **116,** 238-247.

Geshnizjani, N., Sarikhani Khorami, S., Willems, L. A. J., Snoek, B. L., Hilhorst, H. W. M. & Ligterink, W. (2019) The interaction between genotype and maternal nutritional environments affects tomato seed and seedling quality. *J Exp Bot* **70,** 2905-2918.

Geshnizjani, N., Snoek, B. L., Willems, L. A. J., Rienstra, J. A., Nijveen, H., Hilhorst, H. W. M. & Ligterink, W. (2020) Detection of QTLs for genotype x environment interactions in tomato seeds and seedlings. *Plant Cell Environ* . doi:10.1111/pce.13788

Graffelman, J. & van Eeuwijk, F. (2005) Calibration of multivariate scatter plots for exploratory analysis of relations within and between sets of variables in genomic research. *Biometrical Journal* **47,** 863-879.

Gutiérrez, R. A., Lejay, L. V., Dean, A., Chiaromonte, F., Shasha, D. E. & Coruzzi, G. M. (2007) Qualitative network models and genome-wide expression data define carbon/nitrogen-responsive molecular machines in Arabidopsis. *Genome biology* **8,** R7.

He, H., Willems, L. A., Batushansky, A., Fait, A., Hanson, J., Nijveen, H., Hilhorst, H. W. & Bentsink, L. (2016) Effects of parental temperature and nitrate on seed performance are reflected by partly overlapping genetic and metabolic pathways. *Plant and Cell Physiology* **57,** 473-487.

Hochberg, U., Degu, A., Toubiana, D., Gendler, T., Nikoloski, Z., Rachmilevitch, S. & Fait, A. (2013) Metabolite profiling and network analysis reveal coordinated changes in grapevine water stress response. *BMC plant biology* **13,** 184.

Jansen, R. C. & Nap, J.-P. (2001) Genetical genomics: the added value from segregation. *Trends in Genetics* **17,** 388-391.

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

Kaplan, F., Kopka, J., Haskell, D. W., Zhao, W., Schiller, K. C., Gatzke, N., Sung, D. Y. & Guy, C. L. (2004) Exploring the temperature-stress metabolome of Arabidopsis. *Plant physiology* **136,** 4159-4168.

Kazmi, R. H., Khan, N., Willems, L. A., Van Heusden, A. W., Ligterink, W. & Hilhorst, H. W. (2012) Complex genetics controls natural variation among seed quality phenotypes in a recombinant inbred population of an interspecific cross between Solanum lycopersicum× Solanum pimpinellifolium. *Plant, Cell & Environment* **35,**929-951.

Kazmi, R. H., Willems, L. A., Joosen, R. V., Khan, N., Ligterink, W. & Hilhorst, H. W. (2017) Metabolomic analysis of tomato seed germination. *Metabolomics* **13,** 145.

Keurentjes, J. J., Fu, J., De Vos, C. R., Lommen, A., Hall, R. D., Bino, R. J., van der Plas, L. H., Jansen, R. C., Vreugdenhil, D. & Koornneef, M. (2006) The genetics of plant metabolism. *Nature genetics***38,** 842.

Keurentjes, J. J. & Sulpice, R. (2009) The role of natural variation in dissecting genetic regulation of primary metabolism. *Plant signaling & behavior* **4,** 244-246.

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

Khan, N., Kazmi, R. H., Willems, L. A., Van Heusden, A. W., Ligterink, W. & Hilhorst, H. W. (2012) Exploring the natural variation for seedling traits and their link with seed dimensions in tomato.*PLoS One* **7,** e43991.

Kinnersley, A. M. & Turano, F. J. (2000) Gamma aminobutyric acid (GABA) and plant responses to stress. *Critical Reviews in Plant Sciences***19,** 479-509.

Kliebenstein, D. (2009) Advancing genetic theory and application by metabolic quantitative trait loci analysis. *The plant cell***21,** 1637-1646.

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

Kooke, R. & Keurentjes, J. (2011) Multi-dimensional regulation of metabolic networks shaping plant development and performance.*Journal of Experimental Botany* **63,** 3353-3365.

Krasensky, J. & Jonak, C. (2012) Drought, salt, and temperature stress-induced metabolic rearrangements and regulatory networks.*Journal of Experimental Botany* **63,** 1593-1608.

Kumar, N. (2006). *Breeding of horticultural crops: principles and practices* : New India Publishing.

Lehmann, T. & Ratajczak, L. (2008) The pivotal role of glutamate dehydrogenase (GDH) in the mobilization of N and C from storage material to asparagine in germinating seeds of yellow lupine. *Journal of plant physiology* **165,** 149-158.

Li, Y., Breitling, R. & Jansen, R. C. (2008) Generalizing genetical genomics: getting added value from environmental perturbation.*Trends in Genetics* **24,** 518-524.

Li, Y., Swertz, M. A., Vera, G., Fu, J., Breitling, R. & Jansen, R. C. (2009) designGG: an R-package and web tool for the optimal design of genetical genomics experiments. *BMC Bioinformatics* **10,**188.

Lisec, J., Meyer, R. C., Steinfath, M., Redestig, H., Becher, M., Witucka-Wall, H., Fiehn, O., Torjek, O., Selbig, J. & Altmann, T. (2008) Identification of metabolic and biomass QTL in Arabidopsis thaliana in a parallel analysis of RIL and IL populations. *The Plant Journal* **53,** 960-972.

Lommen, A. (2009) MetAlign: interface-driven, versatile metabolomics tool for hyphenated full-scan mass spectrometry data preprocessing.*Analytical chemistry* **81,** 3079-3086.

Michaeli, S., Fait, A., Lagor, K., Nunes-Nesi, A., Grillich, N., Yellin, A., Bar, D., Khan, M., Fernie, A. R. & Turano, F. J. (2011) A mitochondrial GABA permease connects the GABA shunt and the TCA cycle, and is essential for normal carbon metabolism. *The Plant Journal***67,** 485-498.

Morgenthal, K., Weckwerth, W. & Steuer, R. (2006) Metabolomic networks in plants: Transitions from pattern recognition to biological interpretation. *Biosystems* **83,** 108-117.

Nunes-Nesi, A., Fernie, A. R. & Stitt, M. (2010) Metabolic and signaling aspects underpinning the regulation of plant carbon nitrogen interactions. *Molecular plant* **3,** 973-996.

Obata, T. & Fernie, A. R. (2012) The use of metabolomics to dissect plant responses to abiotic stresses. *Cellular and Molecular Life Sciences* **69,** 3225-3243.

Rao, E. S., Kadirvel, P., Symonds, R. C. & Ebert, A. W. (2013) Relationship between survival and yield related traits in Solanum pimpinellifolium under salt stress. *Euphytica* **190,**215-228.

Ratajczak, W., Lehmann, T., Polcyn, W. & Ratajczak, L. (1996) Metabolism of amino acids in germinating yellow lupine seeds. I. The decomposition of 14 C-aspartate and 14 C-glutamate during the imbibition. *Acta physiologiae plantarum* **1** .

Renault, H., El Amrani, A., Palanivelu, R., Updegraff, E. P., Yu, A., Renou, J.-P., Preuss, D., Bouchereau, A. & Deleu, C. (2011) GABA accumulation causes cell elongation defects and a decrease in expression of genes encoding secreted and cell wall-related proteins in Arabidopsis thaliana. *Plant and Cell Physiology* **52,** 894-908.

Rodriguez-Lopez, M., Garzo, E., Bonani, J., Fereres, A., Fernandez-Munoz, R. & Moriones, E. (2011) Whitefly resistance traits derived from the wild tomato Solanum pimpinellifolium affect the preference and feeding behavior of Bemisia tabaci and reduce the spread of Tomato yellow leaf curl virus. *Phytopathology* **101,**1191-1201.

Rosental, L., Nonogaki, H. & Fait, A. (2014) Activation and regulation of primary metabolism during seed germination. *Seed science research* **24,** 1-15.

Rosental, L., Perelman, A., Nevo, N., Toubiana, D., Samani, T., Batushansky, A., Sikron, N., Saranga, Y. & Fait, A. (2016) Environmental and genetic effects on tomato seed metabolic balance and its association with germination vigor. *BMC genomics* **17,**1047.

Schauer, N., Semel, Y., Roessner, U., Gur, A., Balbo, I., Carrari, F., Pleban, T., Perez-Melis, A., Bruedigam, C. & Kopka, J. (2006) Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. *Nature biotechnology***24,** 447.

Shedlarski, J. G. & Gilvarg, C. (1970) The pyruvate-aspartic semialdehyde condensing enzyme of Escherichia coli. *Journal of Biological Chemistry* **245,** 1362-1373.

Shelp, B. J., Bown, A. W. & McLean, M. D. (1999) Metabolism and functions of gamma-aminobutyric acid. *Trends in plant science***4,** 446-452.

Steuer, R., Kurths, J., Fiehn, O. & Weckwerth, W. (2003) Interpreting correlations in metabolomic networks. In: Portland Press Limited.

Stitt, M. & Fernie, A. R. (2003) From measurements of metabolites to metabolomics: an 'on the fly'perspective illustrated by recent studies of carbon–nitrogen interactions. *Current opinion in biotechnology* **14,** 136-144.

Strehmel, N., Hummel, J., Erban, A., Strassburg, K. & Kopka, J. (2008) Retention index thresholds for compound matching in GC–MS metabolite profiling. *Journal of Chromatography B* **871,** 182-190.

Taji, T., Ohsumi, C., Iuchi, S., Seki, M., Kasuga, M., Kobayashi, M., Yamaguchi-Shinozaki, K. & Shinozaki, K. (2002) Important roles of drought-and cold-inducible genes for galactinol synthase in stress tolerance in Arabidopsis thaliana. *The Plant Journal* **29,**417-426.

Tikunov, Y., Laptenok, S., Hall, R., Bovy, A. & De Vos, R. (2012) MSClust: a tool for unsupervised mass spectra extraction of chromatography-mass spectrometry ion-wise aligned data.*Metabolomics* **8,** 714-718.

Toubiana, D., Semel, Y., Tohge, T., Beleggia, R., Cattivelli, L., Rosental, L., Nikoloski, Z., Zamir, D., Fernie, A. R. & Fait, A. (2012) Metabolic profiling of a mapping population exposes new insights in the regulation of seed metabolism and seed, fruit, and plant relations.*PLoS Genetics* **8,** e1002612.

Tschoep, H., Gibon, Y., Carillo, P., Armengaud, P., Szecowka, M., NUNES-NESI, A., Fernie, A. R., Koehl, K. & Stitt, M. (2009) Adjustment of growth and central metabolism to a mild but sustained nitrogen-limitation in Arabidopsis. *Plant, Cell & Environment***32,** 300-318.

Urbanczyk-Wochniak, E. & Fernie, A. R. (2004) Metabolic profiling reveals altered nitrogen nutrient regimes have diverse effects on the metabolism of hydroponically-grown tomato (Solanum lycopersicum) plants.*Journal of Experimental Botany* **56,** 309-321.

Ursem, R., Tikunov, Y., Bovy, A., Van Berloo, R. & Van Eeuwijk, F. (2008) A correlation network approach to metabolic data analysis for tomato fruits. *Euphytica* **161,** 181.

Van Ooijen, J. (2004) MapQTL 5, Software for the mapping of quantitative trait loci in experimental populations. Kyazma BV, Wageningen, the Netherlands. *MapQTL 5, Software for the mapping of quantitative trait loci in experimental populations. Kyazma BV, Wageningen, the Netherlands.* -.

Voorrips, R. E., Verkerke, W., Finkers, R., Jongerius, R. & Kanne, J. (2000) Inheritance of taste components in tomato. *Acta physiologiae plantarum* **22,** 259-261.

Vurro, M., Boari, A., Pilgeram, A. L. & Sands, D. C. (2006) Exogenous amino acids inhibit seed germination and tubercle formation by Orobanche ramosa (broomrape): potential application for management of parasitic weeds. *Biological Control* **36,** 258-265.

Wilson, M. F. & Bell, E. A. (1978) Amino acids and β-aminopropionitrile as inhibitors of seed germination and growth. *Phytochemistry***17,** 403-406.

Windsor, A. J., Reichelt, M., Figuth, A., Svatoš, A., Kroymann, J., Kliebenstein, D. J., Gershenzon, J. & Mitchell-Olds, T. (2005) Geographic and evolutionary diversification of glucosinolates among near relatives of Arabidopsis thaliana (Brassicaceae). *Phytochemistry***66,** 1321-1333.

Table 1. Name, category, *p*.value and Log_2 ratio (HP:LN) of the metabolites in the RILs that significantly changed betwee

Metabolite
Sorbitol
Galactarate
Glycolate
Glycerate
Erythronic acid
Phosphoric acid
Malate
Fructose-6-phospate
Salycilate
Succinate
Ethanolamine
Guanosine
Malonate
Pyroglutamate
Gluconte
Palmitic acid
GABA
Urea
Glutamate
Benzoate
Galactinol
Glycine
Aspartate
Leucine
Glycerol-6-phosphate
2ketoglutaric acid
Fumaric acid

Table 1. Name, category, *p*.value and Log_2 ratio (HP:LN) of the metabolites in the RILs that significantly changed betwee

Citrate
Serine
Trehalose
Phenylalanine
Mannitol
Quinate
Maltose
Threonate
Alanine

*p.value shows the significance level of the metabolite changes by ANOVA analysis between the two maternal environments

Table 2. Properties of the networks constructed from HP and LN seed metabolite levels.

Attributes	**High Phosphate**
Number of nodes	56
Total number of edges	395
Number of positive edges	379
Number of negative edges	16
Range of node degree	1-28
Average number of neighbours	14.11
Network density	0.256

Table 2. Properties of the netv

Table 3. Number of QTLs identified in each sub-population and for the genetic and genotypei × environment component v

Condition
High phosphate (HP)
Low nitrate (LN)
Genetic (G)
Genetic by environment interaction (G×E)

Figure 1. Principal component analysis (PCA) of annotated metabolites in the two parental lines grown in two different nutritional environments. **MM,** *Solanum lycopersicum* (cv. Moneymaker) shown as triangles; **PI,** *Solanum pimpinellifolium* shown as circles; **HP,** High phosphate shown in red; **LN,** Low nitrate shown in green.

Figure 2. A, Principal component analysis (PCA) of known metabolites of dry seeds of RILs grown in **HP,** High phosphate (green circles) and **LN,** Low nitrate (red circles) conditions. **B,** Loading scores of metabolites for PC1 and PC2.**AA,** amino acids (red triangles); **OA,** Organic acids (yellow triangles); **SA,** Sugar alcohols (green triangles);**Others,** Other components (Blue triangle).

Figure 3. Fold change of the metabolites in two nutritional maternal environments. **HP** , High phosphate and **LN** , Low nitrate. **AA** , Amino acids; **OA** , Organic acids;**SA** , Sugar alcohols.

Figure 4. Metabolite profiles in dry tomato seeds. The comparison of metabolite content of the 100 tomato RILs grown in two nutritional maternal environments, **HP** , High phosphate and**LN** , Low nitrate.

Figure 5. Correlation networks of known metabolites for each maternal environment. **A** , Low nitrate; **B** , High phosphate. The colours of the nodes represent the metabolites category.**AA** , Amino acids; **OA** , Organic acids; **SA** , Sugar alcohols. The **Blue** and **Red** colour of the edges (lines) indicate positive and

negative correlations, respectively. The size of the nodes correlates with the number of connections within the network (the degree).

Figure 6. Spearman correlation matrix of all pairs of known metabolites across the whole RIL population derived from *S. lycopersicum* cv. Moneymaker and *S. pimpinellifolium* .

Figure 7. Correlation matrix of metabolites with phenotypic traits within the whole RIL population derived from *S. lycopersicum* cv. Moneymaker and *S. pimpinellifolium* .

Figure 8. Distribution of metabolite variation within the genotypes for each nutritional maternal environment with high phosphate**(HP)** in black and low nitrogen **(LN)** in grey.

Figure 9. Heatmap of LOD profiles of the mQTLs detected for G and G×E. **A,** Heatmap representing the positions of the mQTLs explained by the genetic component (G); **B,** Heatmap indicating the position of the mQTLs affected by genotype by environment interactions (G×E). The 12 chromosomes of tomato are separated by dashed lines. Coloured spots indicating the significant QTLs. The blue and yellow colours show loci where the *S. pimpinellifolium* and the*S.. lycopersicum* alleles enhance the metabolite levels, respectively. Metabolites and their categories are shown at the right side of the panels.