Genes for seed quality

Integrating physiology and genetical genomics to mine for seed quality genes in tomato

Rashid Hussain Kazmi

Thesis Committee

Thesis supervisor

Prof. dr. H.J. Bouwmeester
Professor, Laboratory of Plant Physiology

Thesis co-supervisor
Dr. H.W.M. Hilhorst
Associate Professor, Laboratory of Plant Physiology
Wageningen University

Dr. ir. J.W. Ligterink
Staff Researcher, Laboratory of Plant Physiology
Wageningen University

Other members

Prof. dr. ir. M. Koornneef, Wageningen University and Max Planck Institute for Plant Breeding Research (MPIZ), Cologne, Germany

Prof. dr. B. Finch-Savage, Warwick University, UK

Dr. P. van den Wijngaard, Syngenta, Netherlands

Dr. H.J. Finkers, Wageningen University, Wageningen

This research was conducted under the auspices of the Graduate School of Experimental Plant Sciences

Genes for seed quality

Integrating physiology and genetical genomics to mine for seed quality genes in tomato

Rashid Hussain Kazmi

Thesis

Submitted in fulfillment of the requirements for the degree of doctor

at Wageningen University

by the authority of the Rector Magnificus

Prof. dr. M.J. Kropff,

in the presence of the

Thesis Committee appointed by the Academic Board

to be defended in public

on Wednesday 20 March 2013

at 4 p.m. in the Aula

Rashid H. Kazmi

Genes for seed quality: Integrating physiology and genetical genomics to mine for seed quality genes in tomato

243 pages

Thesis, Wageningen University, Wageningen, NL (2013)

With references, with summaries in Dutch and English

ISBN 978-94-6173-520-1

Contents

Chapter 1
GENERAL INTRODUCTION

1.1 Seed Quality

Seed quality in tomato is thought to be associated with many interlinked physiological and genetical traits (Hilhorst, 2007; Hilhorst, 2010). Seed quality is determined by a number of physiological principles related to important plant developmental processes, such as embryogenesis, growth, stress-resistance and the transition from a seed to an autotrophic seedling (Ouyang et al., 2002; Spanò et al., 2007). Seed quality attributes include germination (percentage, rate and uniformity), dormancy, seed and seedling vigor (germination/growth under stress conditions), seedling dry weight, and normal embryo and seedling morphology, as well as the ability to develop into a normal plant (Goodchild and Walker, 1971; Bewley, 1997; Delseny et al., 2001; Finch-Savage and Leubner-Metzger, 2006; El-Kassaby et al., 2008; Angelovici et al., 2010). Seed germination and seedling growth in tomato are the stages of the plant life that are most sensitive to environmental stresses (e.g. salt, temperature and water loss) (Foolad et al., 1997; Foolad and Chen, 1999; Foolad et al., 2003; Foolad, 2007). These stresses ultimately delay the onset, rate and distribution of the germination events. It is believed that the plant's response to environmental stresses is controlled by many genes (Foolad, 2007). Although recent studies on seed development have been invaluable in revealing aspects of the regulation of metabolism, questions concerning the genetic basis of seed trait variability remain open due to the lack of integrative studies on a population scale. Therefore, there is a need to determine the genetic basis of tomato germination traits under different

stresses. In particular it is imperative to know whether the same or different loci are contributing to seed germination under salt, osmotic, cold, high-temperature and oxidative stress. Seed quality traits, such as seed germination and vigor, as well as protein, starch and oil contents are functionally related to the carbon-nitrogen balance, central metabolism and sink-source interactions during seed development on the mother plant. Post-genomic technologies, such as transcriptomics and metabolomics, are excellent tools for the global analysis of seed/seedling processes associated with quality. The molecular-genetic dissection of these seed processes and their relationship with seed and seedling phenotypes will ultimately identify the regulatory genes and signaling pathways and, thus, provide the means by which to predict and enhance seed quality (Ligterink et al., 2012).

Research to date has tended to focus on the nutritional quality of crops as economic values of seed quality rather than on genetic mechanisms regulating seed quality traits. The issue has grown in importance in the light of proven associations between mother plant traits and seed metabolism and vigor on a genomic scale (Toubiana et al., 2012). The present study seeks to find integrative approaches that could facilitate the understanding of the underlying causes of the complex trait of seed quality. The idea is to provide new ways of dissecting the genetics of seed quality by combining the physiology, genetics and genomics of tomato F_8 Recombinant Inbred Lines (RIL) obtained from a cross between *Solanum lycopersicum (*cv. Moneymaker, 'MM') and *Solanum pimpinellifolium* (G1.1554, 'Pimp') to identify loci and genes that are responsible for seed quality traits. Until now systematic studies to address the seed quality phenomenon in a multidisciplinary way have been lacking. In the present study, we have focused on the systematic exploitation of the naturally

occurring variation in *S. lycopersicum* × *S. pimpinellifolium* F_8 RILs as a complementary resource in order to decipher seed quality phenotypes.

The genetic analysis of natural variation in tomato

Traditionally, genetic variation among accessions has led to the belief that it has limited use, mostly because of its quantitative nature compared to the mutants, which provide qualitative variation (Nadeau and Frankel, 2000; Keurentjes et al., 2008). The existence of differences lies in the number of loci and the environmental effects underlying the variation under study, which determine the tools used for its analysis. However, over the past decade there has been a dramatic increase in efficient molecular marker technologies and specific statistical methods, which has established the map position and the effects of quantitative trait loci (QTL) (Tanksley et al., 1992; Tanksley, 1993; Foolad, 1999; Mackay et al., 2009). There has been an increasing interest in the study of variation among tomato accessions. Several studies have exploited natural variation to answer a lot of questions related to the molecular basis of quantitative traits in tomato (Foolad and Lin, 1998; Foolad et al., 2003) and other crop species, including sunflower, rapeseed and Arabidopsis (Clerkx et al., 2004; Asghari, 2007; Ebrahimi et al., 2008; Bentsink et al., 2010; Perez-Vega et al., 2010; Joosen et al., 2012).

Central to the entire discipline of quantitative genetics is the concept of crosses among various accessions. The resultant progenies derived therefrom segregate for a number of genetic perturbations and can be analyzed genetically for quantitative traits (Keurentjes et al., 2008). In a biological context the association of trait phenotypes with the genotype assayed by molecular markers is very effective for the analysis of QTL. The latter reveals the hotspots on the genetic map containing a locus, or

several closely-linked loci, and their contribution to the total variance of the trait in that experiment. The use of so-called 'immortal mapping populations' consisting of homozygous RILs is an important component in the QTL analysis, and plays a key role in obtaining trait values from different replications and experiments performed in different environments. These are obtained by single-seed descent from F_2 plants until F_6 or further generation(s). These populations especially are of great importance as a number of traits can be mapped in one population. The results of quantitative studies can lead to the discovery that some loci control more than one trait (Koornneef et al., 2004). Co-location of QTLs can also provide a clue to the pathways that might be involved in complex traits.

Complex traits and generalized genetical genomics

One of the most significant current discussions in genome research is about mapping and characterizing trait loci that control variation in various phenotypic characters in plants, e.g. identifying genes that control growth, metabolism, and development. These traits are commonly known as complex traits, and are considered to have a multi-factorial background controlled by an unknown number of QTLs, as well as many environmental factors (Mackay, 2001; Mackay et al., 2009). Contrary to classical Mendelian or monogenic traits, which are controlled by single genes, complex traits usually reflect many small phenotypic contributions of multiple genes (Mackay, 2001; Mackay et al., 2009). The phenotypic modifications may occur due to various molecular modifications, such as single-nucleotide polymorphisms (SNPs) or small and large sequence deletions in the coding regions, or in the regulatory non-coding regions, that influence protein levels and/or function (Foolad, 1996; Mackay, 2001; Glazier et al., 2002; Mackay et al., 2009).

Dissection of the molecular basis, even for a monogenic trait, is complicated since each gene, or its protein product, interacts with many other genes, proteins and pathways. Thus, the gap between genotype and phenotype remains enormous and indeed the identification of the functional mutation and molecular basis of complex traits has only been successful for a very small proportion of QTLs. Recent work has begun to shift the methodology for better understanding of complex traits to a more promising approach of genetical genomics. The genetical genomics approach integrates traditional QTL mapping with gene expression and metabolic profiling studies for a better understanding of the mechanisms influencing complex traits (Joosen et al., 2009; Ligterink et al., 2012). Although, genetical genomic studies takes the effect of genetic perturbations on biological systems at the molecular level, it usually does not take into account the environmental conditions. Thus, a comprehensive understanding of biological systems requires studying them across multiple environments as the molecular networks largely depend on environmental cues. The present study seeks to address this knowledge gap by using a generalized-genetical-genomics (GGG) approach (Li et al., 2008) for tomato seed transcriptomic and metabolomic analysis. Our GGG approach takes into account genetics and chosen environmental perturbations (different seed developmental stages, i.e. dry and imbibed seeds) in combination with the analysis of the genetic variation present in F_8 RILs to identify genotype-by-environment interactions. Hence, the application of a GGG model, which is essentially a systems genetics approach, provides a broad overview of changes in expression and primary metabolic processes that occur during dry and imbibed tomato seed developmental stages. Thus, the present approach reveals, for the first time, the plasticity of molecular networks in tomato for seed quality traits and forms a crucial step toward

understanding different influences of genetic and developmental responses of tomato seeds.

Transcriptomics and metabolomics for the dissection of complex traits

Technological advances have redefined what we call 'phenotype': in the past a phenotype was a one dimensional entity but the combination of morphological, transcriptional, protein and metabolic readouts associated with a particular combination of alleles has turned it into a multi-dimensional entity. Nonetheless, "omics" technologies have enabled further integration of different data sets. They assume that the gene expression levels are also affected by the functional polymorphism that affects the trait of interest (Arbilly et al., 2006). In genetical genomics, the expression level of each transcript is treated as a quantitative phenotype and the marker genotypes are used to map loci affecting the gene expression levels, known as expression QTL (eQTL) (Jansen and Nap, 2001). The idea is to use segregation and recombination of related individuals (e.g. RILs, ILs, NILs); then, for example, each individual of the population is used for genetic mapping and gene expression analysis. Thus, gene expression values of all the individuals in a segregating population are treated as a quantitative trait for QTL mapping. The eQTLs can be described as 'cis-acting', when the eQTL is located in the same postion as the gene that is affected, and as 'trans-acting' when the eQTL and affected gene are not in the same genomic position (Jansen, 2003). Cis-acting eQTLs are found to be more significant as they have larger effects on transcription and can be considered as key positional candidates for the functional QTL (Mata et al., 2005; Keurentjes et al., 2007; West et al., 2007). One of the first approaches to integrate QTL studies with gene expression profiling was in

yeast (Brem et al., 2002), soon followed by organisms like eucalyptus (Kirst et al., 2004), maize (Schadt et al., 2003) and Arabidopsis (Keurentjes et al., 2007). Most studies using genetical-genomics have been carried out in *A. thaliana*, mainly due to availability of high quality mapping populations and the commercially available genome-wide micro-arrays. Several studies in various RIL populations have indicated extensive genetic regulation of gene expression (Keurentjes et al., 2007; West et al., 2007; Cubillos et al., 2012).

The most significant goal in metabolomics is the comprehensive measurement of, as many as possible metabolites in a sample. A large and growing body of literature has investigated the fact that metabolite abundance is generally controlled by multiple genes and are thus quantitative in nature and regulated by metabolite QTL (mQTL) (Keurentjes et al., 2008; Lisec et al., 2008; Schauer et al., 2008; Toubiana et al., 2012). Combinatorial broad-spectrum and quantitative genetic analysis of metabolites in plants allowed the detailed molecular dissection of metabolic biosynthetic pathways, as well as QTL mapping of plant metabolites (Magrath et al., 1993; Kliebenstein et al., 2001; Kliebenstein et al., 2001; Keurentjes et al., 2007; Meyer et al., 2007; Rowe et al., 2008). An integral link exists between plant central metabolism and development/physiology, which is largely based upon empirical studies that investigate relationships of metabolites and developmental variation (Keurentjes et al., 2007; Meyer et al., 2007). However, in many instances QTLs for metabolite and developmental traits did not co-locate more than was expected by chance. A variety of reasons have been suggested for this lack of overlap between known developmental and metabolite QTLs, including the size of the structured mapping populations (Beavis, 1998; Clerkx et al., 2004; Rowe et al., 2008). In turn this gives rise to the view that genetic regulation of plant

metabolism is more complex than presumed, such that current studies lack sufficient power to detect the majority of metabolite QTLs present in a population.

Despite QTL mapping's long success in detecting genetic regions responsible for phenotypic traits, the identification of the functional mutation and molecular basis of complex traits has only been successful in a very small proportion of QTL (Magrath et al., 1993; Flint-Garcia et al., 2003; Mackay et al., 2009). An example is a study in tomato, where the cause of a seed weight QTL has been pinpointed to a gene encoding an ABC transporter gene by using genetic analysis (Orsi and Tanksley, 2009). Bentsink *et al.*, (2010) compared the dry seed transcriptomes of near isogenic lines (NILs) for 'Delay of Germination' (DOG) QTLs of Arabidopsis that differ in after-ripening and/or dormancy, and unraveled genetic and molecular pathways controlling variation for these traits. Another promising example of finding a causal gene by exploiting natural variation was demonstrated in QTL mapping studies of a lettuce RIL population for thermotolerance (Argyris et al., 2005; Argyris et al., 2008). The authors mapped the Htg6.1 QTL in this RIL population and this QTL was further validated in NILs where it was subsequently confirmed to extend the range of germination thermotolerance. Further mapping of candidate genes resulted in the identification of 9-cis-epoxycarotenoid dioxygenase 4 (*LsNCED4*) gene as causal for the Htg6.1 locus controlling germination under high temperature (Schwember and Bradford, 2010). A causal role for reactive oxygen species (ROS) in sunflower embryo dormancy release was also proposed, based on genetic analysis (Oracz et al., 2007).

Networks

The great expectation from genetical genomics is the potential to reconstruct gene expression networks and enabling the integration of data from multiple sources (e.g. genotypic, molecular, metabolomic and expression profiling) (Papp et al., 2011). These integrative approaches could facilitate the understanding of the underlying causes of complex traits of seed quality. Another promising expectation is that the GGG model could lead to a better understanding of environmental and stress responses of seeds as part of integrative seed quality traits, including seed performance under osmotic-, salt-, cold- and heat stress. Metabolomic and expression data from tomato seeds that are exposed to different developmental conditions could show how tomato seeds respond to different abiotic and developmental environments. This concept was applied, and even expanded, by combining the results of gene expression analysis with QTL mapping of metabolite levels and enzyme activities (Keurentjes et al., 2008; Keurentjes and Sulpice, 2009; Keurentjes, 2009). Common genetic-map positions of differentially expressed genes and QTLs allow the construction of genetic networks (Keurentjes et al., 2007; Meyer et al., 2007). A combined analysis of this information within the context of a systems framework holds promise for the future.

QTL confirmation and fine mapping

Quantitative genetics has enjoyed a renaissance in the past decade. Although substantial progress has been made in cloning QTL genes and reducing some of them to Quantitative Trait Nucleotides (QTNs), QTL mapping remains a challenging task as it produces large genetic intervals as well as QTLs of large effect which can split into multiple QTLs, explaining only a small proportion of the total variance. High-precision

mapping leading to fine mapping and cloning of gene(s) responsible for complex trait variation usually requires confirmation and validation of QTLs in different populations. Near-isogenic lines (NILs) that differ for markers flanking the QTL analysis facilitate the confirmation of QTLs and ultimately the identification of genes underlying complex traits (Robertson et al., 1988; Kaeppler et al., 1993; Balasubramanian et al., 2009; Kooke et al., 2012). The NIL analysis allows dissection of QTL into smaller intervals as they differ in respect of overlapping regions of the genome indicated by QTL analysis (Tuinstra et al., 1997). Alternatively, NILs contrasting at QTLs can be developed by selection within heterogeneous inbred families (HIFs) (Tuinstra et al., 1997). HIFs are a set of lines derived from RILs that are genetically similar but have not reached complete homozygosity and segregate for those loci that were heterozygous. Molecular markers can be used to screen a population of HIFs derived from different inbreds to identify families that segregate for a specific region of the genome (Tuinstra et al., 1997). This approach can be used to develop a series of NILs that contrast for a specific genomic region. This HIF concept is effective because one does not have to create the NILs first, which requires several generations of backcrossing and marker-assisted selection (MAS). Using either NILs and/or HIFs allows the validation/confirmation of the presence and the effect of a QTL. NILs have a genetic background consisting of only one genotype and HIFs have a mixed allelic (but identical) background. To locate the genomic position of a QTL more precisely, lines with recombinant events in the QTL region can be phenotyped. The lines that show the expected effect according to the QTL detection/validation should carry the gene responsible for the QTL effect. Thus, a subset of RILs used to develop HIFs further delineates the QTLs interval and allows the validation, and thus further fine mapping. In the

fine mapping approach, lines with recombinant events in the QTL region are phenotyped and the correlation of phenotype and genotype narrows down the region of interest, ideally to the underlying gene. Transcriptome analysis can be used in parallel with a fine mapping approach to identify differentially expressed genes in the region of interest as possible candidate genes for a QTL.

Thesis Motivation

There is currently a great interest in the investigation of systematic methods for the dissection of complex traits governing seed quality. Most of these studies were driven by the interest of chasing regulatory genes, which might control complex networks and were able to identify causal genes responsible for a certain trait or disease (Secko, 2005). Arabidopsis studies have assayed several thousand eQTLs of large phenotypic effects, but almost all (93%) of the 36,871 eQTLs were associated with small phenotypic effects. Many transcripts/e-traits were controlled by multiple eQTLs with opposite allelic effects and exhibited higher heritability in the RILs than their parents, suggesting non-additive genetic variation. To our knowledge, this is the first large-scale global eQTL study in a relatively large plant mapping population (West et al., 2007). It reveals that the genetic control of transcript level is highly variable and multifaceted and that this complexity may be a general characteristic of eukaryotes (West et al., 2007). Another example is the genome-wide gene expression in the parents and a RIL population of a cross between the Landsberg *erecta* (L*er*) and Cape Verde Islands (Cvi). Transcript levels of 24,065 genes were analyzed by DNA micro-arrays, of which 922 showed significant differential expression between the parents. Subsequent mapping resulted in 4,523 eQTLs detected for 4,066 genes (Keurentjes et al., 2007). Similarly, a study

in Eucalyptus analyzed 2,608 genes of an *E. grandis* × *E. globules* backcross population to reveal the genetic networks responsible for growth variation. Two loci were discovered to control lignin biosynthesis coordinately. In addition, these two loci were localized in growth-related QTLs. The authors suggested that the targeted regions might regulate growth, lignin content and composition (Kirst et al., 2004). Some of these genetical genomics initial findings made the field very popular. However, the exploration and integration of the available data originating from the various experimental areas, has not, as yet, been achieved. In order to exploit the data and make it more interpretable and useful for the evaluation of seed quality phenotypes, a systematic way is needed to integrate and analyze the results generated by quantitative trait analyses, microarray and metabolomic studies and molecular biology.

Thesis objective

The objective of this thesis is to explore molecular–genetic methods, tools and frameworks in order to obtain a better understanding of the mechanisms governing complex traits of tomato seed quality. The aim is to be able to characterize identified QTLs in the best possible way; (1) to ascertain which loci are likely to be responsible for a certain trait; (2) how these loci interact with each other; (3) what is the proportion in which the environment (non-stress vs. stress) affects the phenotypic traits; (4) what are the downstream (global) effects; (5) which loci have been reported in the same region as that in which the identified QTL was reported. This thesis attempts to get closer to the biological molecular-genetic interpretation of high-throughput data and the genetic characterization of QTLs by exploiting various sources of information and

bringing them together, and ultimately to target potential candidate genes that could be responsible for a certain seed quality phenotypic trait.

Chapter 2 introduces the concepts of QTL mapping and looks at how the genetic variation present in a tomato RIL population controls the regulation of different germination indices. This chapter also presents a review of the co-locating QTLs for germination under non-stress and stress conditions indicating the genetic relationships between germination phenotypes under stress and non-stress environments and subsequent possibilities for improvement of tomato seed germination using selection.

In **Chapter 3** morphological assessments of seedlings are made. One of the more significant aspects of this study is its emphasis on seed dimensions, such as seed size and seed length, which was ignored. A strong relationship between different seed/seedling dimensions and root architecture was found, cementing the argument that larger food reserves in large-sized seed helps in establishing a more extensive root system.

Chapter 4 assesses the systems-genetics approach to find link between primary metabolites and seed quality phenotypes. The concept of generalized genetical genomics (GGG) with environmental perturbations (different seed developmental stages, i.e. dry and imbibed seeds) in combination with the analysis of genetic variation for metabolite abundance present in the RIL population is comprehended. The large-scale genetic analysis of metabolite abundance has clearly shown their usefulness in predicting germination phenotypes and constructing correlation network modules, hence narrowing the QTL-phenotype gap.

Chapter 5 deals with post QTL analyses and methods exemplified through the construction of the HIFs for confirmation/validation of the QTLs. HIFs were constructed using the heterozygousity present in the F_8 of the *S. lycopersicum* x *S. pimpinellifolium* RIL population that allowed the

unequivocal recognition of QTL regions along chromosomes 6 and 8 of the tomato genome.

Chapter 6 features the concluding remarks of the thesis and a final critical opinion about present and future research needed to follow up for the analyses of complex traits.

References:

Angelovici R, Galili G, Fernie AR, Fait A (2010) Seed desiccation: a bridge between maturation and germination. Trends in Plant Science **15:** 211-218

Arbilly M, Pisante A, Devor M, Darvasi A (2006) An integrative approach for the identification of quantitative trait loci. Animal genetics **37:** 7-9

Argyris J, Dahal P, Hayashi E, Still DW, Bradford KJ (2008) Genetic variation for lettuce seed thermoinhibition is associated with temperature-sensitive expression of abscisic acid, gibberellin, and ethylene biosynthesis, metabolism, and response genes. Plant Physiology **148:** 926-947

Argyris J, Truco MJ, Ochoa O, Knapp SJ, Still DW, Lenssen GM, Schut JW, Michelmore RW, Bradford KJ (2005) Quantitative trait loci associated with seed and seedling traits in Lactuca. TAG Theoretical and Applied Genetics **111:** 1365-1376

Asghari A (2007) QTL analysis for cold resistance-related traits in Brassica napus using RAPD markers. International journal of food, agriculture and environment **5:** 188-192

Balasubramanian S, Schwartz C, Singh A, Warthmann N, Kim MC, Maloof JN, Loudet O, Trainer GT, Dabi T, Borevitz JO (2009) QTL mapping in new Arabidopsis thaliana advanced intercross-recombinant inbred lines. PLoS ONE **4:** e4318

Beavis WD (1998) QTL analyses: power, precision, and accuracy. Molecular dissection of complex traits**:** 145-162

Bentsink L, Hanson J, Hanhart CJ, Blankestijn-de Vries H, Coltrane C, Keizer P, El-Lithy M, Alonso-Blanco C, de Andres MT, Reymond M, van Eeuwijk F, Smeekens S, Koornneef M (2010) Natural variation for seed dormancy in Arabidopsis is regulated by additive genetic and molecular pathways. Proceedings of the National Academy of Sciences of the United States of America **107:** 4264-4269

Bewley JD (1997) Seed germination and dormancy. Plant Cell **9:** 1055-1066

Brem RB, Yvert G, Clinton R, Kruglyak L (2002) Genetic dissection of transcriptional regulation in budding yeast. Science **296:** 752-755

Clerkx EJM, El-Lithy ME, Vierling E, Ruys GJ, Blankestijin-De Vries H, Groot SPC, Vreugdenhil D, Koornneef M (2004) Analysis of natural allelic variation of Arabidopsis seed germination and seed longevity traits between the accessions Landsberg erecta and Shakdara, using a new recombinant inbred line population. Plant Physiology **135:** 432-443

Cubillos FA, Yansouni J, Khalili H, Balzergue S, Elftieh S, Martin-Magniette ML, Serrand Y, Lepiniec L, Baud S, Dubreucq B (2012) Expression variation in connected recombinant populations of Arabidopsis thaliana highlights distinct transcriptome architectures. BMC genomics **13:** 117

Delseny M, Bies-Etheve N, Carles C, Hull G, Vicient C, Raynal M, Grellet F, Aspart L (2001) Late embryogenesis abundant (LEA) protein gene regulation during Arabidopsis seed maturation. Journal of Plant Physiology **158:** 419-427

Ebrahimi A, Maury P, Berger M, Kiani SP, Nabipour A, Shariati F, Grieu P, Sarrafi A (2008) QTL mapping of seed-quality traits in sunflower recombinant inbred lines under different water regimes. Genome **51:** 599-615

El-Kassaby YA, Moss I, Kolotelo D, Stoehr M (2008) Seed germination: Mathematical representation and parameters extraction. Forest Science **54:** 220-227

Finch-Savage WE, Leubner-Metzger G (2006) Seed dormancy and the control of germination. New Phytologist **171:** 501-523

Flint-Garcia SA, Thornsberry JM, IV B (2003) Structure of Linkage Disequilibrium in Plants*. Annual Review of Plant Biology **54:** 357-374

Foolad MR (1996) Unilateral incompatibility as a major cause of skewed segregation in the cross between Lycopersicon esculentum and L-pennellii. Plant Cell Reports **15:** 627-633

Foolad MR (1999) Comparison of salt tolerance during seed germination and vegetative growth in tomato by QTL mapping. Genome **42:** 727-734

Foolad MR (2007) Genome mapping and molecular breeding of tomato. International Journal of Plant Genomics **2007:** 52

Foolad MR, Chen FQ (1999) RFLP mapping of QTLs conferring salt tolerance during the vegetative stage in tomato. Theoretical and Applied Genetics **99:** 235-243

Foolad MR, Lin GY (1998) Genetic analysis of low-temperature tolerance during germination in tomato, Lycopersicon esculentum Mill. Plant Breeding **117:** 171-176

Foolad MR, Stoltz T, Dervinis C, Rodriguez RL, Jones RA (1997) Mapping QTLs conferring salt tolerance during germination in tomato by selective genotyping. Molecular Breeding **3:** 269-277

Foolad MR, Subbiah P, Kramer C, Hargrave G, Lin GY (2003) Genetic relationships among cold, salt and drought tolerance during seed germination in an interspecific cross of tomato. Euphytica **130:** 199-206

Glazier AM, Nadeau JH, Aitman TJ (2002) Finding genes that underlie complex traits. Science **298:** 2345-2349

Goodchild NA, Walker MG (1971) A method of measuring seed germination in physiological studies. Annals of Botany **35:** 615-621

Hilhorst H.W.M. F-SWE, Buitink J., Bolingue W. & Leubner-Metzger G. (2010) Dormancy in plant seeds. In Dormancy and Resistance in Harsh Environments (eds E. Lubzens, J. Cerda & M. Clark), pp. 43–67. Springer, Berlin, Germany.

Hilhorst HWM (2007) Definition and hypotheses of seed dormancy. In: K.J. Bradford and H. Nonogaki (Eds.) Chapter 4: Seed development, dormancy and germination. Annual Plant Reviews Vol. 27. Blackwell Publishing, Sheffield, UK**:** 50-71

Jansen RC (2003) Studying complex biological systems using multifactorial perturbation. Nature Reviews Genetics **4:** 145-151

Jansen RC, Nap JP (2001) Genetical genomics: the added value from segregation. TRENDS in Genetics **17:** 388-390

Joosen R, Ligterink W, Hilhorst H, Keurentjes J (2009) Advances in genetical genomics of plants. Current Genomics **10:** 540

Joosen RVL, Arends D, Willems LAJ, Ligterink W, Jansen RC, Hilhorst HWM (2012) Visualizing the genetic landscape of Arabidopsis seed performance. Plant Physiology **158:** 570-589

Kaeppler S, Phillips R, Kim T (1993) Use of near-isogenic lines derived by backcrossing or selfing to map qualitative traits. TAG Theoretical and Applied Genetics **87:** 233-237

Keurentjes J, Sulpice R (2009) The role of natural variation in dissecting genetic regulation of primary metabolism. Plant Signaling & Behavior **4:** 244

Keurentjes J, Sulpice R, Gibon Y, Steinhauser M, 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

Keurentjes JJB (2009) Genetical metabolomics: closing in on phenotypes. Current Opinion in Plant Biology **12:** 223-230

Keurentjes JJB, Bentsink L, Alonso-Blanco C, Hanhart CJ, Vries HBD, Effgen S, Vreugdenhil D, Koornneef M (2007) Development of a near-isogenic line population of Arabidopsis thaliana and comparison of mapping power with a recombinant inbred line population. Genetics **175:** 891-905

Keurentjes JJB, Fu J, Terpstra IR, Garcia JM, Van Den Ackerveken G, Snoek LB, Peeters AJM, Vreugdenhil D, Koornneef M, Jansen RC (2007) Regulatory network construction in Arabidopsis by using genome-wide gene expression quantitative trait loci. Proceedings of the National Academy of Sciences **104:** 1708-1713

Keurentjes JJB, Sulpice R, Gibon Y, Steinhauser MC, 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

Kirst M, Myburg AA, De León JPG, Kirst ME, Scott J, Sederoff R (2004) Coordinated genetic regulation of growth and lignin revealed by quantitative trait locus analysis of cDNA microarray data in an interspecific backcross of eucalyptus. Plant Physiology **135:** 2368-2378

Kliebenstein DJ, Gershenzon J, Mitchell-Olds T (2001) Comparative quantitative trait loci mapping of aliphatic, indolic and benzylic

glucosinolate production in Arabidopsis thaliana leaves and seeds. Genetics **159:** 359-370

Kliebenstein DJ, Lambrix VM, Reichelt M, Gershenzon J, Mitchell-Olds T (2001) Gene duplication in the diversification of secondary metabolism: Tandem 2-oxoglutarate–dependent dioxygenases control glucosinolate biosynthesis in Arabidopsis. The Plant Cell Online **13:** 681-693

Kooke R, Wijnker E, Keurentjes J (2012) Backcross populations and near isogenic lines. Methods in molecular biology (Clifton, NJ) **871:** 3

Koornneef M, Alonso-Blanco C, Vreugdenhil D (2004) Naturally occurring genetic variation in Arabidopsis thaliana. Annual Review of Plant Biology **55:** 141-172

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

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

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

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

Mackay TFC (2001) The genetic architecture of quantitative traits. Annual Review of Genetics **35:** 303-339

Mackay TFC, Stone EA, Ayroles JF (2009) The genetics of quantitative traits: challenges and prospects. Nature Reviews Genetics **10:** 565-577

Magrath R, Herron C, Giamoustaris A, Mithen R (1993) The inheritance of aliphatic glucosinolates in Brassica napus. Plant breeding **111:** 55-72

Mata J, Marguerat S, Bahler A (2005) Post-transcriptional control of gene expression: a genome-wide perspective. Trends in Biochemical Sciences **30:** 506-514

Meyer RC, Steinfath M, Lisec J, Becher M, Witucka-Wall H, Torjek O, Fiehn O, Eckardt A, Willmitzer L, Selbig J, Altmann T (2007) The metabolic signature related to high plant growth rate in Arabidopsis thaliana. Proceedings of the National Academy of Sciences of the United States of America **104:** 4759-4764

Nadeau JH, Frankel WN (2000) The roads from phenotypic variation to gene discovery: mutagenesis versus QTLs. Nature genetics **25:** 381-384

Oracz K, Bouteau HEM, Farrant JM, Cooper K, Belghazi M, Job C, Job D, Corbineau F, Bailly C (2007) ROS production and protein oxidation as a novel mechanism for seed dormancy alleviation. The Plant Journal **50:** 452-465

Orsi CH, Tanksley SD (2009) Natural variation in an ABC transporter gene associated with seed size evolution in tomato species. PLoS genetics **5:** e1000347

Ouyang X, van Voorthuysen T, Toorop PE, Hilhorst HWM (2002) Seed vigor, aging, and osmopriming affect anion and sugar leakage during imbibition of maize (Zea mays L.) caryopses. International Journal of Plant Sciences **163:** 107-112

Papp B, Notebaart RA, Pál C (2011) Systems-biology approaches for predicting genomic evolution. Nature Reviews Genetics **12:** 591-602

Perez-Vega E, Paneda A, Rodriguez-Suarez C, Campa A, Giraldez R, Ferreira JJ (2010) Mapping of QTLs for morpho-agronomic and seed quality traits in a RIL population of common bean (Phaseolus vulgaris L.). Theoretical and Applied Genetics **120:** 1367-1380

Robertson D, Victor B, Helfman G, Schultz E, Warner R, Searcy W, Pleszczynska W, Hansell R, Lili A, Hollman S (1988) Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature **335:** 20

Rowe HC, Hansen BG, Halkier BA, Kliebenstein DJ (2008) Biochemical networks and epistasis shape the Arabidopsis thaliana metabolome. Plant Cell **20:** 1199-1216

Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V, Ruff TG, Milligan SB, Lamb JR, Cavet G (2003) Genetics of gene expression surveyed in maize, mouse and man. Nature **422:** 297-302

Schauer N, Semel Y, Balbo I, Steinfath M, Repsilber D, Selbig J, Pleban T, Zamir D, Fernie AR (2008) Mode of inheritance of primary metabolic traits in tomato. Plant Cell **20:** 509-523

Schwember AR, Bradford KJ (2010) A genetic locus and gene expression patterns associated with the priming effect on lettuce seed germination at elevated temperatures. Plant Molecular Biology **73:** 105-118

Secko D (2005) Genetics embraces expression. The Scientist **19:** 26-27

Spanò C, Buselli R, Ruffini Castiglione M, Bottega S, Grilli I (2007) RNases and nucleases in embryos and endosperms from naturally aged wheat seeds stored in different conditions. Journal of Plant Physiology **164:** 487-495

Tanksley SD (1993) Mapping polygenes. Annual Review of Genetics **27:** 205-233

Tanksley SD, Ganal MW, Prince JP, Devicente MC, Bonierbale MW, Broun P, Fulton TM, Giovannoni JJ, Grandillo S, Martin GB, Messeguer R, Miller JC, Miller L, Paterson AH, Pineda O, Roder MS, Wing RA, Wu W, Young ND (1992) High-density molecular linkage maps of the tomato and potato genomes. Genetics **132:** 1141-1160

Toubiana D, Semel Y, Tohge T, Beleggia R, Cattivelli L, Rosental L, Nikoloski Z, Zamir D, Fernie AR, 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

Tuinstra M, Ejeta G, Goldsbrough P (1997) Heterogeneous inbred family (HIF) analysis: a method for developing near-isogenic lines that differ at quantitative trait loci. TAG Theoretical and Applied Genetics **95:** 1005-1011

West MAL, Kim K, Kliebenstein DJ, Van Leeuwen H, Michelmore RW, Doerge R, Clair DAS (2007) Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in Arabidopsis. Genetics **175:** 1441-1450

Chapter 2

SEED QUALITY PHENOTYPES IN A RECOMBINANT INBRED POPULATION OF AN INTERSPECIFIC CROSS BETWEEN *SOLANUM LYCOPERSICUM* X *SOLANUM PIMPINELLIFOLIUM*

Published in Plant Cell & Environment
Kazmi RH*, Khan N*, Willems LAJ, van Heusden AW, Ligterink W, Hilhorst HWM (2012), Vol. 35: 929-951

*Equal contribution

Abstract

Seed quality in tomato is associated with many complex physiological and genetic traits. While plant processes are frequently controlled by the action of small to large-effect genes that follow classic Mendelian inheritance, our study suggests that seed quality is primarily quantitative and genetically complex. Using a recombinant inbred line population of *Solanum lycopersicum* x*Solanum pimpinellifolium*,we identified quantitative trait loci influencing seed quality phenotypes under non-stress, as well as salt-, osmotic-, cold-, high temperature- and oxidative stress conditions. In total 42 seed quality traits were analyzed and 120 QTLs were identified for germination traits under different conditions. Significant phenotypic correlations were observed between germination traits under optimal conditions, as well as under different stress conditions. In conclusion, one or more QTLs were identified for each trait with some of these QTLs co-locating. Co-location of QTLs for different

traits can be an indication that a locus has pleiotropic effects on multiple traits due to a common mechanistic basis. However, several QTLs also dissected seed quality in its separate components, suggesting different physiological mechanisms and signaling pathways for different seed quality attributes.

Introduction

Seed quality is the ability of seeds to germinate under a wide variety of environmental conditions and to develop into healthy seedlings. Seed quality is determined by several factors including genetic and physical purity, mechanical damage and physiological conditions, such as viability, germination, dormancy, vigor and uniformity (Dickson, 1980; Hilhorst and Toorop, 1997; Hilhorst, 2007; Hilhorst et al., 2010). The physiological condition of seeds during development and maturation has a strong effect on ultimate seed quality. It is influenced by several environmental factors such as temperature, humidity, light and nutrients during the seed filling and maturation stages, by seed treatments (harvesting and processing) and by accumulated damage (Ouyang et al., 2002; Spano et al., 2007). Thus, seed quality is a complex trait governed by interactions between the genome and the environment (Koornneef et al., 2002) and therefore, seed quality can be challenged over the entire seed production chain. These quality-specific interactions are primarily expressed as germination, which is defined as the event that begins with the uptake of water by the seed and ends with the start of elongation by the embryonic axis, usually the protrusion of the radicle (Bewley, 1997; Finch-Savage and Leubner-Metzger, 2006). In the case of tomato, protrusion of the radicle through the surrounding layers (endosperm and testa) is considered to be the

completion of germination. Thus, successful germination is determined by the balance between two opposing forces.

Abiotic stresses, such as extreme temperatures, low water availability, high salt levels, mineral deficiency and toxicity, are frequently encountered by plants in both natural and agricultural systems (Langridge et al., 2006; Eswaran et al., 2010). Higher plants have developed strategies to avoid abiotic stresses whereas these strategies are lost in agricultural crops. The most striking effect of abiotic stresses is on the yield of crops, which is estimated to be less than half under abiotic stress, as compared to normal growing conditions. Traditional approaches to improve the abiotic stress tolerance of crop plants by breeding have been of very limited success. This is mainly because of the difficulty of selecting for stress tolerance traits in traditional breeding programs. However, the natural variation among crop species can be used to cross desired traits from wild relatives and, for tomato, extensive abiotic stress tolerance has been identified in screens of land races and related wild species. Nevertheless, there is relatively little known about the molecular basis of abiotic stress tolerance in tomato species and there is still ample scope for improvement.

Substantial genetic variation for abiotic stresses exists within the cultivated tomato (*Solanum lycopersicum;* Wudiri and Henderson, 1985; Moyle and Muir, 2010), as well as in its related wild species, such as *Solanum habrochaitis, Solanum pimpinellifolium*, and *Solanum pennellii*. These wild species offer the genetic resources for cold, temperature, and water stress tolerance with respect to seed quality (Foolad and Lin, 1998; Foolad et al., 2003). However, rather limited efforts have been devoted to the physiological and genetic characterization of this variation in tomato to warrant its use for developing drought-tolerant cultivars (Kahn et al., 1993; Martin et al., 1999). This is in contrast with the considerable amount of

research that has been conducted on abiotic stress in relation to other crop species, including rice (*Oryza sativa* L.; Zhang et al., 2001) and lettuce; Johnson et al., 2000). In a recent germplasm evaluation study, several wild tomato cultivars were identified as possessing the ability to germinate rapidly under abiotic stresses, including *S. pimpinellifolium* Mill. accession LA722 (Foolad et al., 2003). *S. lycopersicum* is sensitive to cold-, salt- and drought stress during seed germination, whereas *S. pimpinellifolium* germinates rapidly under most conditions, including cold-, salt-, and drought stress. Among the wild species of tomato, S*. pimpinellifolium* is the most closely related to *S. lycopersicum* and the only species for which natural introgression with *S. lycopersicum* has been demonstrated (Rick, 1958). Accessions within this species are red fruited and can be readily hybridized with the cultivated tomato. Furthermore, in comparison with other wild tomato species, *S. pimpinellifolium* possesses fewer undesirable horticultural characteristics and thus has been frequently used as a genetic resource in tomato genetics and breeding programs.

Crop performance is the end result of the action of thousands of genes and their interaction with the environment. Conventional breeding has been very successful in raising the yield potential of crops (Borlaug and Dowswell, 2003; Campos et al., 2004; Collins et al., 2008). Breeders have exploited genetic variability for crop improvement with very limited knowledge of factors governing it. However, this approach may become inadequate as the pressure to provide improvements will mount if global climate change increases the frequency and severity of abiotic constraints. Temperature stress, drought and salinity will be more prevalent in marginal areas with an increased demand for agricultural products and reduced availability of arable land and natural resources, such as water and fertilizers. Consequently, the genetic dissection of the quantitative traits

controlling the adaptive response of crops to abiotic stress is a prerequisite to allow cost-effective applications of genomics-based approaches to breeding programs aimed at improving the sustainability and stability of yield under adverse conditions.

Consistent with the proposition that seed quality has a complex genetic basis, QTL studies of seed quality have generally revealed the influence of numerous QTLs of small to large phenotypic effect. Quantitative trait mapping of seed quality traits in common bean, sunflower, rapeseed, tomato and Arabidopsis has revealed numerous QTLs (Foolad et al., 2003; Foolad et al., 2003; Clerkx et al., 2004; Asghari, 2007; Ebrahimi et al., 2008; Bentsink et al., 2010; Perez-Vega et al., 2010). *S. lycopersicum* is severely susceptible to environmental stresses (e.g. salt, drought, cold and high temperature) during seed germination and seedling growth, delaying the onset, rate and distribution of the germination events (Foolad et al., 2007). To take up the challenges manifested in uncovering the causal polymorphisms for QTLs, genomics tools are now also available for *S. lycopersicum* and these offer promising opportunities to unravel network mechanisms underlying complex quantitative traits (Collins et al., 2008). To elucidate the molecular mechanisms underlying quantitative traits, we analyzed quantitative responses of tomato seed quality phenotypes in a structured RIL mapping population.

In the present study we used the recombinant inbred line population generated from *S. lycopersicum* (cv. Moneymaker) and *S. pimpinellifolium* (G1.1554) (Voorrips et al., 2000). This population provides a valuable resource for the study of genes affecting complex phenotypes for seed quality as they allow isolation of the effect of a particular QTL from those of the entire genome, thus increasing our statistical power to dissect

quantitative seed quality phenotypes, shaping a complex underlying mechanism.

Materials and Methods

Plant Material

Solanum lycopersicum cv. Moneymaker, a horticulturally superior, advanced tomato breeding line, was crossed with *Solanum pimpinellifolium* G1.1554 , a self-compatible inbred accession of the wild species to produce 83 recombinant inbred lines (RILs) to F_8 (Voorrips et al., 2000). This population was genotyped for a total of 865 SNP markers in F_7.

Growth Conditions and Seed Collection

The *Solanum lycopersicum x Solanum pimpinellifolium* RIL population was grown twice under controlled conditions in the greenhouse facilities at Wageningen University, The Netherlands. The day and night temperatures were maintained at 25 °C and 15 °C, respectively, with 16 hours light and 8 hours dark (long-day conditions). All the RILs were uniformly supplied with the basic dose of fertilizers and other nutrients. Seeds were extracted from healthy fruits and treated with 1% hydrochloric acid (HCL) to remove the large pieces of the pulp sticking onto the seeds. The solution of tomato seed extract with diluted hydrochloric acid was passed through a fine mesh sieve and washed with water to remove the remaining parts of the pulp and remnants of the hydrochloric acid. The seeds were processed and disinfected by soaking in a solution of tri-sodium phosphate ($Na_3PO_4.12H_2O$). Finally, seeds were dried on clean filter paper at room temperature and were brushed to remove impurities with a seed brusher (Seed Processing Holland BV, www.seedprocessing.nl). The cleaned seeds were dried for 3 days at 20 °C and were stored in a cool, dry storage room (13 °C and 30% RH) in paper bags.

Linkage Analysis

The genetic linkage map consists of 12 individual linkage groups corresponding to the 12 chromosomes of tomato. Sequence information was used to study the segregation of parental alleles in the *Solanum lycopersicum* G1.1554 x *Solanum pimpinellifolium* cv. Moneymaker Recombinant inbred lines (RIL) population. Custom made Infinium Bead arrays, containing 5529 Single Nucleotide Polymorphisms (SNP), were used to genotype the RIL population. In total 5529 SNP markers were used to genotype *S. pimpinellifolium* G1.1554 and *S. lycopersicum* cv Moneymaker. The identical markers (no recombination between two markers) were removed and left 2251 polymorphic markers out of 5529 SNPs. The loci with identical segregation patterns were removed before calculating the map. The remaining 865 unique markers were used for calculating the maps of all chromosomes. Map construction was done in JoinMap 4 (Van Ooijen and Voorrips, 2001) based on recombination frequency and Haldane's mapping function by incorporating the available SNP marker data set for 83 RILs. The name of each marker on the tomato linkage map corresponds to the position on the tomato genome sequence version SL2.40 (http://solgenomics.net/organism/solanum_lycopersicum/genome).

Seed phenotyping

Germination assay

Germination assays were performed in triplicate with seeds of the parents and the RILs, which were sown under aseptic conditions on germination trays (21×15 cm DBP Plastics, http://www.dbp.be) containing 15 ml water (non-stress condition) or NaCl, polyethylene glycol (PEG) or H_2O_2 (stress-conditions), and one layer of white filter paper (20.2 x 14.3 cm white blotter paper; Allpaper, http://www.allpaper.nl). Each germination tray contained 2 lines and 45 seeds of each line and was considered one

replicate. Germination trays were placed in a completely randomized design with three replications per sample. A maximum of 17 trays were piled up with two empty trays on both the top and the bottom end of the stack, with 15 ml water and two layers of white filter paper, to prevent unequal evaporation. The trays were covered with tightly fitting lids and the whole pile was wrapped in a closed transparent plastic bag and incubated at 4 °C for 3d for stratification. Subsequently the bags where placed randomly in an incubator at 25 °C in the dark (type 5042; Seed Processing Holland, http://www. seedprocessing.nl), except for brief intervals when germination was counted under laboratory (fluorescent) lighting. Germination responses were scored visually as radicle protrusion at 8-hourly intervals for 10 consecutive days during the period of most rapid germination, and at longer subsequent intervals, until no additional germination was observed.

Salt, osmotic and oxidative stress

Salt, osmotic and oxidative stress tolerance treatments were applied in germination trays with 15 ml of the corresponding solution on a piece of filter paper. Salt stress was estimated by germinating seeds in different concentrations of NaCl. Osmotic potentials were established through aqueous solutions of polyethylene glycol (PEG 8000, Sigma) measured in mega Pascal (MPa). Specific concentrations of NaCl and PEG 8000 were determined with the Solute Potential and Molar-Molal-g Solute/g Water Interconversion (SPMM) program (Michel and Radcliffe, 1995). Tolerance to hydrogen peroxide was estimated by germinating seeds on filter paper saturated with a solution of 300 mM H_2O_2.

Low- and high-temperature stress

All RIL genotypes were subjected to sub-optimal temperature regimes in order to test their response to temperature stress. Germination was monitored during incubation for 10 days at 12 °C in the case of cold stress, and at 35 °C and 36 ºC to test for high-temperature stress response.

Statistical and genetic analyses

Calculation of G_{max}, t_{10}^{-1}, t_{50}^{-1}, MGR, U_{7525}^{-1}, AUC and estimation of means

In this study the curve-fitter module of the Germinator package was used for analyzing different parameters of the cumulative germination curves (Joosen et al., 2010). Parental lines and the RIL population were subjected to different germination conditions, and maximum germination (G_{max}, %), the onset of germination (t_{10}^{-1}; reciprocal of time to 10 % of germination of viable seeds (h^{-1})), the rate of germination (t_{50}^{-1}; reciprocal of time to 50 % of the germination of viable seeds (h^{-1}), MGR = mean germination rate, which is reciprocal of the mean germination time (MGT^{-1}), uniformity (U_{7525}^{-1}, reciprocal of time interval between 75 and 25% viable seeds to germinate; h^{-1}), and area under the germination curve (AUC; the integration of the fitted curve between t = 0 and a user-defined endpoint (x)) were determined. A full description of the validity and assessment of calculated parameters is available elsewhere (Thomson and El-Kassaby, 1993; Bradford, 1995; Hayashi et al., 2008; Alonso-Blanco et al., 2009; Landjeva et al., 2010). The t_{10}^{-1}, t_{50}^{-1} and U_{7525}^{-1} were calculated only for those treatments where seeds of the majority of RILs (>80 %) completed a corresponding fraction (10, 50, 75% or more) of germination (Hayashi et al., 2008; Galpaz and Reymond, 2010). For germination parameters the mean of the three replicates were calculated and these were transformed

to a probit regression model using the R module "VGAM" (http://www.r-project.org). Means of transformed data were used for QTL analysis.

Identification of QTLs

QTL analysis was performed on the basis of the established marker linkage map of the RIL population, which contains 865 SNP markers. The mapping software MapQTL®5.0 (Van Ooijen and Maliepaard, 2003) was used for identifying QTL positions in the genome for a given trait. A multiple QTL mapping model (MQM) was used to identify potential QTLs (Jansen et al., 1995) as implemented in MapQTL®5.0. In this method, background markers are selected to take over the role of the putative QTL as co-factors to reduce the residual variance. A two-stage MQM analysis was performed. In the first stage ,conventional interval mapping was performed at a 2 cM interval; the LOD profiles from interval mapping were inspected and the marker closest to each LOD peak was selected as the co-factor to perform further MQM mapping analysis. Several cycles were performed to obtain the potentially maximum number of co-factors for the MQM analysis. These co-factor markers were then subjected to backward elimination, as implemented in MapQTL®5.0, in order to select the best model for the second stage MQM analysis. Such a backward elimination procedure leaves out one co-factor at a time in order to create a subset of co-factors. The likelihood of each of these subset models is compared with the likelihood of the full model with all co-factors, and the subset model which causes the smallest change in likelihood is chosen as the starting set for a subsequent round of elimination. This process continues until the change in likelihood is significant according to the0.002 *P*-value for the test. The set of co-factors then retained was used in the second stage of the MQM analysis. In the final LOD profile, QTLs were affirmed according

to the threshold LOD scores ranging from 2.0 to 7.0 (genome-wide false-positive rate 5%), depending on chromosome map length and the number of chromosome pairs (Van Ooijen, 1999). To determine whether QTLs among different traits were significantly co-located, first, the number of QTLs from different traits that had overlapping confidence intervals were determined. Then, QTL confidence intervals were randomized across the genome 1,000 times, and the distribution of the number of overlapping QTLs of different traits determined. If this number of randomized QTLs was less than the original QTL overlap 95% of the time, the co-location was deemed significant.

Analysis of heritability and epistasis

Broad-sense heritability (h^2_b) was estimated from one-way random-effects of analysis of the variance (ANOVA, SPSS version 19.0) with the equation:

$h^2_b = \sigma^2_g / (\sigma^2_g + \sigma^2_e)$

where σ^2_g is the genetic variance and σ^2_e is the environmental variance (Keurentjes et al., 2007). Significant differences among all means of the RILs were estimated using one-way ANOVA followed by a least significant difference (LSD) test.

A two-dimensional genome-wide epistatic interactions analysis was performed using the R/qtl software package (Broman et al. 2003) in order to identify epistatic interactions contributing to variation in the seed germination parameters: G_{max}, t_{10}^{-1}, t_{50}^{-1}, MGR, U_{7525}^{-1} and AUC. Each chromosomal region (tomato chromosomes 1–12) was considered jointly with all other chromosomal regions throughout the genome for each seed quality phenotype analyzed. The statistical analysis of epistasis as implemented in the R/qtl software package consists of nested linear model-fitting for each pair of loci tested for an epistatic interaction, as described

previously (Koller et al., 2009). To obtain appropriate genome-wide significance thresholds for the epistasis results and properly account for the large number of tests considered in the genome-by-genome scan, 10,000 permutation tests (Doerge and Churchill, 1996) were performed with the Haley-Knott regression method (Broman et al., 2003). In this manner the LOD significance threshold of the maximum genome-wide interaction was found to be 4.09; for full model (lod.full), and conditional interactive model (lod.fv) LOD significance thresholds were found to be 6.04 and 4.63, respectively. Interacting QTL pairs were only reported if all of these thresholds were exceeded. Specifically, the 42 traits measured of each recombinant inbred line were randomly reassigned as a group across the 83 RILs resulting in a permuted data set (Spano et al., 2007). By keeping all phenotypic data together, the underlying phenotypic correlations were preserved. The epistasis analysis was then performed across the whole genome and the resulting maximum LOD scores for linkage for each phenotype were recorded.

Results

Distribution, means and heritability

To investigate the genetic architecture of seed quality traits, we measured phenotypes of the 83 F_8 RILs. The population was derived from a cross between *Solanum lycopersicum* (cv. Moneymaker) and *Solanum pimpinellifolium* (G1.1554). Seeds of the wild accession *S. pimpinellifolium* G1.1554 germinated significantly more rapidly than seeds of the breeding line *S. lycopersicum* cv. Moneymaker under non-stress (control) as well as salt-, osmotic-, cold-, and temperature stress conditions (Table 2.1). The germination parameters were calculated only for those traits in which a corresponding fraction (10, 50, 75% or more) of seeds completed

germination. For example, undercontrol- and salt- (-0.3MPa NaCl) seeds from the majority of RILs surpassed 80% of germination and all parameters like G_{max}, t_{10}^{-1}, t_{50}^{-1}, MGR, U_{7525}^{-1} and AUC were obtained. On the other hand if final germination fell below the particular fraction, t_{10}^{-1} and t_{50}^{-1}, those traits were not calculated, for example, in case of osmotic- (-0.3, -0.5 MPa PEG), cold- (12 °C), high-temperature- (36 °C) and oxidative stress conditions, G_{max}, t_{10}^{-1}, MGR and AUC were obtained but t_{50}^{-1} and U_{7525}^{-1} were not, as the final germination percentage was too low to calculate meaningful values.

In most cases seeds of the RIL population germinated intermediately between the two parental lines, indicating the inheritance of rapid germination from G1.1554 to the progeny (Table 2.1, Figure 2.1). However, we also observed transgressive segregation for the seed quality traits (Table 2.1, Figure 2.1). This implies that the different seed phenotypes shown in the *S. lycopersicum and S. pimpinellifolium* parental lines result from the presence of distinct genetic polymorphisms with antagonistic effects contributed by each parent. Estimates of the broad-sense heritability of different seed quality traits differed considerably among seed phenotypes studied across different treatments (Table 2.2). Heritability estimates for different germination-related traits indicated that genetic variation exists for seed quality phenotypes under control conditions, as well as salt-, osmotic-, cold-, high temperature- and oxidative stress conditions and the germination characteristics in the RIL population are highly heritable (Table 2.1). The RILs showed great phenotypic variation with regard to seed quality traits; G_{max} showed a slight negative skew and t_{10}^{-1}, t_{50}^{-1}, MGR, U_{7525}^{-1} and AUC a stronger positive skew (Figure 2.1).

Table 2.1. Means of germination traits (±SD) for the parental genotypes and the F_8population of cross between *S. lycopersicum* (Money) and *S. pimpinellifolium* (Pimp) in the control- (non-stress), salt-, osmotic-, cold-, temperature- and oxidative-stress treatments.

Treatment	Genotypes	G_{max}	$t_{10}^{-1}{}_{(x100)}$	$t_{50}^{-1}{}_{(x100)}$	$MGR_{(x100)}$	$U_{7525}^{-1}{}_{(x100)}$	AUC
Control	Money	100.0±0.0	1.703±0.032	1.237±0.054	1.198±0.061	3.865±0.488	115.3±4.9
	Pimp	100.0±0.0	3.663±0.106	2.910±0.005	2.652±0.061	17.762±0.290	165.0±0.4
	RILs	92.5±11.3	2.390±0.682	1.811±0.559	1.799±0.607	7.111±4.168	127.0±27.5
Salt I (-0.3MPa NaCl)							
	Money	85.1±1.0	1.230±0.048	0.960±0.040	0.954±0.035	3.823±0.156	77.7±1.8
	Pimp	99.6±0.4	2.609±0.209	2.016±0.159	1.974±0.155	7.682±0.476	148.0±3.4
	RILs	86.7±16.1	1.547±0.419	1.180±0.319	1.170±0.301	4.804±2.400	94.5±30.3
Salt II (-0.5MPa NaCl)							
	Money	85.7±0.8	0.694±0.030	0.502±0.019	0.498±0.001	nd	16.7±4.1
	Pimp	99.6±0.4	1.659±0.101	1.234±0.002	1.200±0.013	nd	115.9±1.0
	RILs	67.9±29.6	1.153±0.392	0.857±0.278	0.840±0.262	nd	57.1±39.7
Osmotic I (-0.3MPa PEG)							
	Money	46.9±19.4	0.810±0.096	nd	0.653±0.051	nd	14.0±6.4
	Pimp	95.5±2.9	1.594±0.256	nd	1.107±0.162	nd	102.5±15.8
	RILs	54.7±28.7	1.176±0.470	nd	0.844±0.261	nd	43.9±15.6
Osmotic II (-0.5MPa PEG)							
	Money	38.3±9.4	0.629±0.046	nd	0.563±0.002	nd	8.31±1.5
	Pimp	70.8±6.3	0.872±0.061	nd	0.698±0.045	nd	28.9±6.5
	RILs	57.8±19.5	0.773±0.202	nd	0.638±0.099	nd	20.2±10.4
Cold Stress 12 °C							
	Money	5.2±2.2	nd	nd	nd	nd	nd
	Pimp	100.0±0.0	0.853±0.048	nd	0.754±0.025	nd	68.5±3.9
	RILs	37.2±18.3	0.568±0.125	nd	0.508±0.080	nd	9.5±3.3
High Temperature I (35 °C)							
	Money	72.8±8.2	1.224±0.130	0.736±0.122	0.751±0.087	nd	45.5±9.7
	Pimp	100.0±0.0	2.803±0.012	2.426±0.009	2.305±0.003	nd	158.2±0.1
	RILs	77.6±28.1	1.889±0.695	1.359±0.510	1.325±0.507	nd	93.2±35.6
High Temperature II (36 °C)							
	Money	3.1±1.3	nd	nd	nd	nd	nd
	Pimp	93.1±3.6	2.507±0.226	nd	1.788±0.139	nd	134.0±8.9
	RILs	33.9±15.9	1.826±0.764	nd	1.254±0.416	nd	39.0±14.5
Oxidative Stress (300mM H_2O_2)							
	Money	64.2±2.7	0.796±0.032	nd	0.642±0.013	nd	24.3±4.8
	Pimp	3.1±0.9	nd	nd	nd	nd	nd
	RILs	0.4±19.4	0.816±0.281	nd	0.649±0.124	nd	17.8±9.6

Money, *Solanum lycopersicum*; Pimp, *Solanum pimpinellifolium*; G_{max} (%), maximum germination; t_{10}^{-1}, t_{50}^{-1}, reciprocal of time to respectively 10 and 50% of viable seeds to germinate (h^{-1}); MGR, mean germination rate (reciprocal of the mean germination time; MGT^{-1}); U_{7525}^{-1},uniformity (reciprocal of time interval between 75 and 25% viable seeds to germinate; h^{-1}); AUC, area under the germination curve (integration of fitted curve between 0 and 200 h); nd, not determined; RIL, recombinant inbred line.

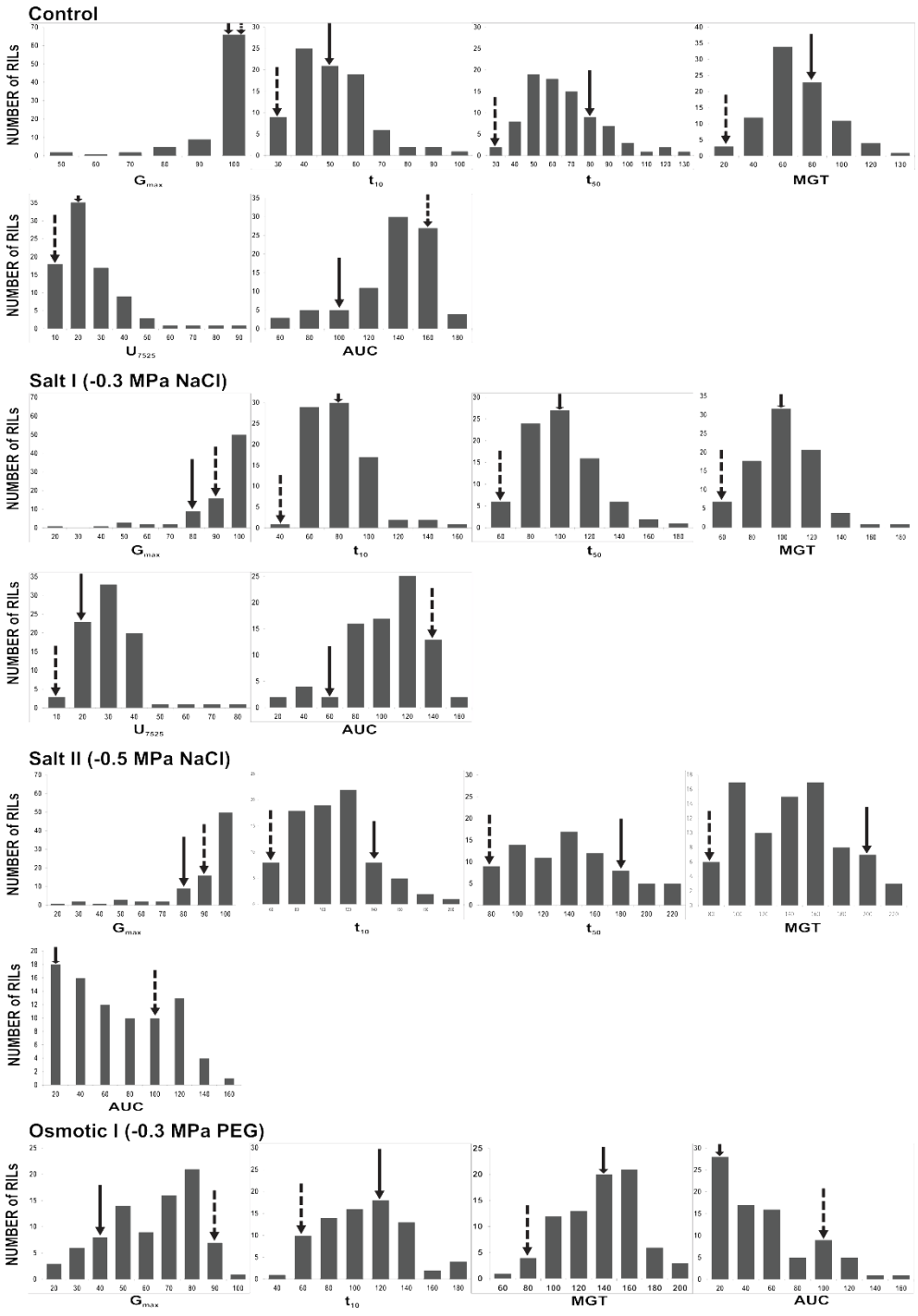


Figure 2.1. *Continued*

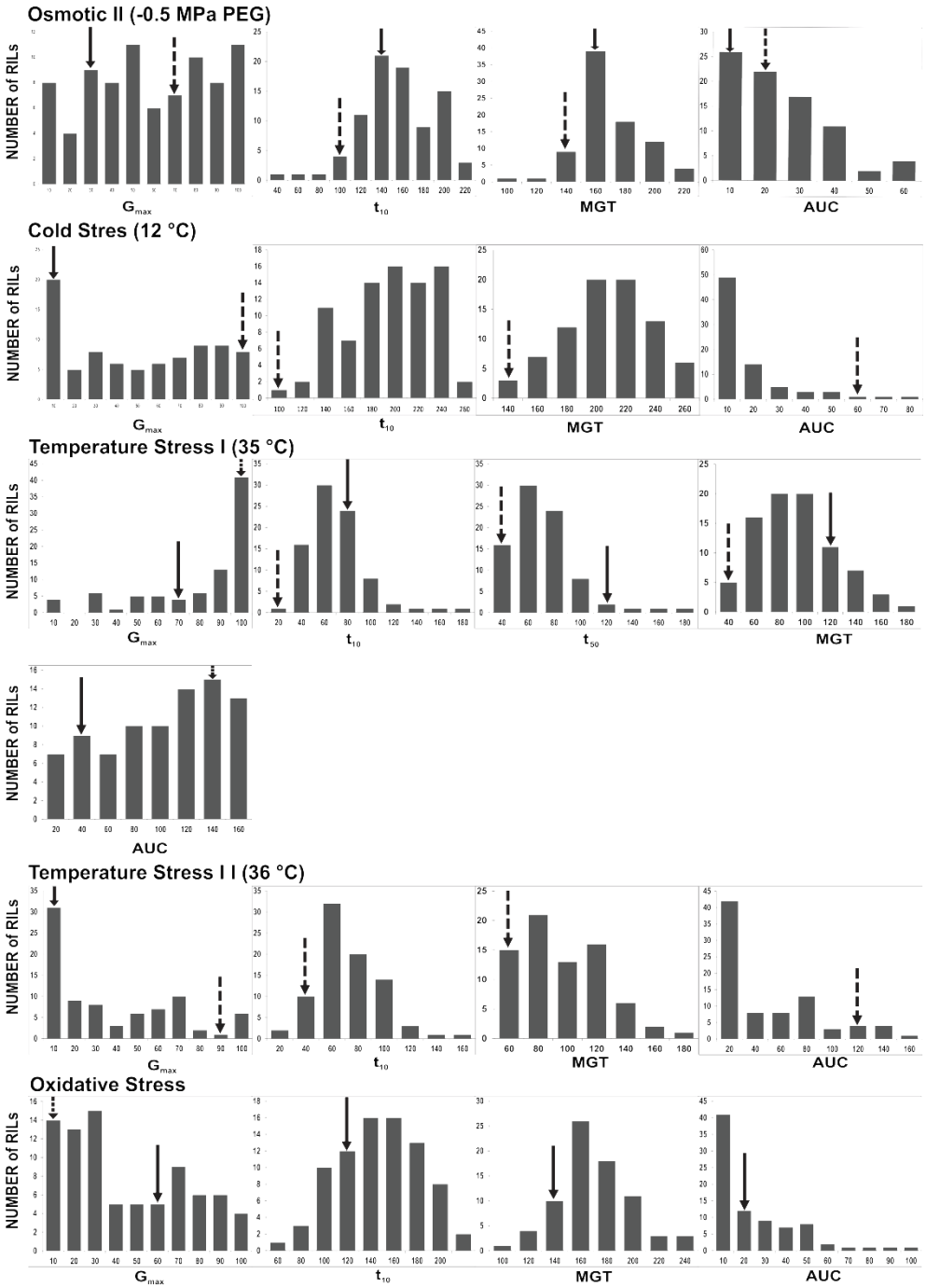


Figure 2.1. Frequency distributions of non-normalized data of all traits in the *Solanum lycopersicum* and *Solanum pimpinellifolium* recombinant

inbred line (RIL) population. Seed quality traits determined under control conditions, salt stress I (-0.3 MPa NaCl), salt stress II (-0.5 MPa NaCl), osmotic stress I (-0.3 MPa PEG), osmotic stress II (-0.5 MPa PEG), cold stress (12 °C), high-temperature stress I (35 °C), high-temperature stress II (36 °C) and oxidative stress.The average parental value is indicated with a solid arrow for *S. lycopersicum* and a dashed arrow for *S. pimpinellifolium* parents. AUC, area under the germination curve; MGT, mean germination time.

Identification of QTLs for germination potential under different conditions

The map position and characteristics of the QTLs associated with the studied seed phenotypes under non-stress (control) and stress-conditions are summarized in Table 2.2 and Table 2.3. We found that individual QTLs mapped to specific regions of the tomato genome. We used an LOD threshold of 2.0 to investigate putative QTLs where seed quality phenotypes map. Figure 2.2 displays a heatmap of LOD profiles. In this way QTLs can be visualized and global 'hot spots' and empty regions across the 12 chromosomes can be seen (Figure 2.2).

QTL for germination under non-stress conditions

To distinguish between loci specific for regulation of germination traits under stress versus non-stress conditions, the latter were determined using the germination traits i.e. G_{max}, t_{10}^{-1}, t_{50}^{-1}, MGR, U_{7525} and AUC. The germination phenotypes were calculated only for those traits in which a corresponding fraction (10, 50, 75% or more) of seeds completed germination. Although we did analyze rate of germination using a number of rate traits (t_{10}^{-1}, t_{50}^{-1}, MGR) as stated in Table 2.1, in order to avoid

repetition and unnecessary complication, we will explicitly discuss t_{10}^{-1} in the results. One QTL was detected for G_{max} on chromosome 7 with an explained variance of 11.9 % (Table 2.2 and 2.3, Figure 2.2). QTL analysis revealed five loci for t_{10}^{-1}, one each on chromosomes 4, 9, 12 and two on chromosome 6. In total these loci accounted for 59.3% of explained variance (Table 2.2 and 2.3, Figure 2.2). Four QTLs were identified for U_{7525}^{-1} on chromosomes 3, 4, 7 and 8, which explained 42.9% of the total variance observed. Two QTLs were revealed for AUC one each on chromosome 2 and 4 which explained 22.6% of the total variance (Table 2.2 and 2.3, Figure 2.2).

QTL for germination under salt stress conditions

Several QTLs were found to be associated with the tested germination traits (Tables 2.2 and 2.3) at -0.3 MPa (low) and -0.5 MPa (high) NaCl levels. For G_{max} one QTL was found on chromosomes 5 at -0.3 MPa and two QTLs were revealed at -0.5 MPa one each on chromosomes 4 and 5, which explained 15.7 and 27.1% of the total variance observed, respectively (Table 2.2 and 2.3, Figure 2.2). For t_{10}^{-1} four QTLs were found one each on chromosomes 4, 6, 11, and 12 under -0.3 MPa which explained variance of 39.3%, whereas three loci were revealed on chromosome 2, 4 and 6 at -0.5 MPa with a total explained variance of 22.7% (Table 2.2 and 2.3, Figure 2.2). Furthermore, for U_{7525}^{-1} under low salt stress two QTLs were identified at chromosomes 4 and 7. In total these loci explained 22.2% of the variance, whereas in the case of high salt level U_{7525}^{-1} was not calculated as the as the majority of RILs did not reach a final germination percentage above 75 %. For AUC two QTLs were found on chromosome 6 and 11 at -0.3 MPa which explained 26.1% of the variance and three QTLs were revealed for -0.5 MPa NaCl on chromosome

Table 2.2. Chromosomal location of the QTL associated with seed quality traits of tomato *Solanum lycopersicum/Solanum pimpinellifolium* RIL population under control (non-stress), salt, osmotic, cold, high-temperature and oxidative stress conditions.

	Chr[a].	Marker Peak[b]	Support Intervalc (cM)	LOD[d]	Explained Variance[e] (%)	Total Explained Variance[f] (%)	Effects[g]	Heritability[h]
Control								
G_{max}								0.89
	7	1559291	0.0-26.0	2.28	11.9	11.9	1.39	
t_{10}^{-1}								0.88
	4	8777285	73.1-80.4	3.39	11.8	59.3	0.70	
	6	34100828	38.9-56.2	3.42	9.8		0.64	
	6	43582592	102.7-108.3	5.59	20.6		0.96	
	9	66917748	100.8-112.7	2.78	9.4		-0.66	
	12	47845308	41.4-64.0	2.30	7.7		-0.57	
t_{50}^{-1}								0.89
	4	56570524	65.1-80.4	2.29	7.7	36.8	0.29	
	6	43582592	101.1-107.3	3.74	13.1		0.37	
	8	57099504	72.6-87.8	2.25	7.6		-0.29	
	12	47845308	49.5-63.0	2.48	8.4		-0.30	
MGR								0.90
	4	8777285	65.1-81.9	2.15	7.2	36.6	0.57	
	6	43582592	101.1-107.3	3.68	12.9		0.75	
	8	57099504	72.6-87.8	2.32	7.9		-0.59	
	12	47845308	49.5-63.0	2.54	8.6		0.61	
U_{7525}^{-1}								0.92
	3	58802824	71.7-82.6	3.34	12.8	42.9	-0.72	
	4	59678612	86.9-108.3	3.23	12.3		0.72	
	7	28075704	33.7-56.7	2.6	9.6		-0.69	
	8	57099504	72.5-86.6	2.25	8.2		-0.81	
AUC								0.80
	2	34914156	23.7-34.2	2.57	12.3	22.6	0.96	
	4	56475308	69.1-81.9	2.18	10.3		0.69	
Salt I(-0.3MPa NaCl)								
G_{max}								0.93
	5	6711122	59.8-66.4	3.32	15.7	15.7	-0.42	

Table 2.1 *Continued*

	Chr[a].	Marker Peak[b]	Support Intervalc (cM)	LOD[d]	Explained Variance[e] (%)	Total Explained Variance[f] (%)	Effects[g]	Heritability[h]
t_{10}^{-1}								0.79
	4	56475308	65.1-87.0	2.27	8.1	39.3	0.58	
	6	43582592	99.5-109.3	3.2	11.7		0.71	
	11	5472482	10.7-17.8	3.07	11.2		-0.70	
	12	44987792	48.9-54.5	2.33	8.3		-0.60	
t_{50}^{-1}								0.94
	4	56475308	68.2-85.0	2.77	8.5	52.1	0.30	
	6	43582592	101.1-109.3	4.89	15.9		0.42	
	9	66917748	106.5-112.7	2.13	6.4		-0.27	
	11	47008280	20.7-36.3	2.61	8		-0.29	
	12	44987792	48.9-54.5	4.18	13.3		-0.38	
MGR								0.89
	1	7044030	51.8-65.7	2.02	5.6	57.7	-0.50	
	4	57013608	68.2-85.0	2.94	9.3		0.62	
	6	43582592	102.4-108.3	5.15	17.4		0.90	
	9	66917748	106.5-112.7	2.41	7.5		-0.59	
	11	5472482	9.0-17.8	2.19	6.8		-0.54	
	12	44987792	48.9-54.5	3.45	11.1		-0.69	
U_{7525}^{-1}								0.94
	4	1767382	0.0-18.9	2.01	9.0	22.2	0.65	
	7	28075704	39.2-56.3	2.88	13.2		0.75	
AUC								0.86
	6	44674784	100.5-112.7	3.16	13.7	26.1	0.81	
	11	48283252	22.7-35.3	2.86	12.4		-0.74	
Salt II(-0.5MPa NaCl)								
G_{max}								0.85
	4	58174884	85.0-93.2	3.13	14.4	27.1	0.79	
	5	7533961	60.7-67.8	2.78	12.7		-0.37	
t_{10}^{-1}								0.68
	2	33752308	4.6-26.7	2.14	7.2	22.7	0.59	
	4	58081284	73.1-95.1	2.26	7.7		0.85	
	6	43763060	99.5-112.7	2.3	7.8		0.54	

Table 2.1 *Continued*

	Chr[a].	Marker Peak[b]	Support Intervalc (cM)	LOD[d]	Explained Variance[e] (%)	Total Explained Variance[f] (%)	Effects[g]	Heritability[h]
t_{50}^{-1}								0.79
	4	58081284	85.0-93.2	2.95	11.4	31.0	0.35	
	6	43763060	99.5-112.7	2.99	11.6		0.35	
	8	57099504	78.4-84.8	2.11	8.0		-0.29	
MGR								0.85
	2	33752308	15.4-26.6	3.43	15.7	15.7	1.04	
AUC								0.72
	4	58174884	85.0-93.2	3.36	12.5	33.9	0.78	
	6	43582592	101.1-109.3	3.56	13.4		0.82	
	9	66917748	99.6-112.7	2.20	8.0		-0.64	
Osmotic I(-0.3MPa PEG)								
G_{max}								0.91
	4	58174884	74.1-93.2	2.62	11.2	29.9	0.70	
	5	6711122	55.7-67.8	2.38	10.1		-0.32	
	9	48774	0.00-12.1	2.05	8.6		0.60	
t_{10}^{-1}								0.85
	2	31348124	7.6-22.4	3.57	16.8	27.2	0.94	
	4	4654114	41.0-52.8	2.28	10.4		0.65	
MGR								0.89
	2	33752308	7.6-23.7	3.5	14.8	34.8	0.96	
	4	4711015	41.0-54.1	2.83	11.8		0.72	
	12	7536683	39.4-64.0	2.02	8.2		-0.61	
AUC								0.85
	4	58174884	64.1-95.1	3.53	12.8	48.1	0.74	
	6	43702064	102.4-108.3	4.94	18.7		0.94	
	9	66917748	106.5-112.7	2.7	9.6		-0.68	
	12	4397607	35.2-48.3	2.01	7.0		-0.56	
Osmotic II(-0.5MPa PEG)								
G_{max}								0.87
	4	54541392	64.1-78.8	2.22	11.9	11.9	0.72	
t_{10}^{-1}								0.83
	2	34914156	9.6-31.3	2.88	7.9	62.3	0.77	
	4	59678612	93.4-100.0	4.95	14.4		0.81	

Table 2.1 *Continued*

	Chr[a].	Marker Peak[b]	Support Intervalc (cM)	LOD[d]	Explained Variance[e] (%)	Total Explained Variance[f] (%)	Effects[g]	Heritability[h]
	6	43023484	101.1-107.3	5.16	15.2		0.83	
	9	66260384	98.9-105.3	2.98	8.2		-0.62	
AUC								0.88
	2	34914156	26.7-33.3	3.45	11.7	44.1	0.94	
	4	54541392	61.2-80.4	2.95	9.8		0.65	
	12	47976208	57.2-62.4	5.58	16.6		-0.86	
MGR								0.53
	2	33752308	15.4-31.3	2.94	13.5	13.5	0.91	
	6	43046416	99.5-108.3	3.07	10.3		0.70	
	12	47976208	57.2-63.0	3.60	12.3		-0.75	
Cold Stress (12 °C)								
G_{max}								0.88
	1	69227784	54.7-65.6	2.07	9.6	32.4	-0.61	
	5	2166131	9.6-35.0	2.89	14.0		0.72	
	6	44674784	105.3-112.0	3.39	8.8		0.78	
t_{10}^{-1}								0.65
	4	4935940	27.4-54.1	2.38	11.9	42.1	0.69	
	5	2515287	20.6-38.6	2.08	9.1		0.62	
	6	43582592	101.1-109.3	2.24	11.2		0.67	
	7	1559291	7.0-38.7	2.29	9.9		-0.64	
MGR								0.74
	7	3317484	24.0-42.2	2.63	12.6	12.6	-0.79	
AUC								
	1	69227784	61.7-65.7	2.4	8.9	37.0	-0.61	0.86
	3	57499392	53.0-76.7	2.01	7.5		-0.55	
	6	44674784	99.5-112.7	2.53	10.7		0.71	
	11	48586064	13.3-35.3	2.33	9.9		-0.63	
High Temperature I (35 °C)								
G_{max}	11	46408368	18.7-30.0	2.86	14.7	14.7	-0.37	0.91
t_{10}^{-1}								0.8
	4	55076292	65.1-72.5	2.49	10.0	33.0	0.64	
	6	44674784	101.1-112.1	5.31	23.0		1.02	

Table 2.1 *Continued*

	Chr[a].	Marker Peak[b]	Support Intervalc (cM)	LOD[d]	Explained Variance[e] (%)	Total Explained Variance[f] (%)	Effects[g]	Heritability[h]
t_{50}^{-1}								0.79
	6	44674784	101.1-112.1	4.38	19.5		0.48	
MGR								0.88
	1	69227784	61.7-65.7	2.79	11.6	35.0	-0.71	
	6	44674784	101.1-112.1	4.63	23.4		1.04	
AUC								0.9
	4	58340636	85.0-96.1	2.01	7.6	30.7	0.58	
	6	43763060	101.1-110.1	3.82	15.2		0.85	
	11	46408368	18.7-32.3	2.08	7.9		-0.58	
High Temperature Stress II (36 °C)								
G_{max}								0.89
	6	43582592	97.5-109.3	2.39	12.7	12.7	0.70	
t_{10}^{-1}								0.75
	6	44905196	110.1-112.7	5.36	31.6	42.5	1.16	
	9	66710096	101.5-112.7	2.09	10.9		-0.69	
MGR								0.93
	6	44905196	111.1-112.7	4.89	28.8	41.7	1.10	
	9	66710096	103.5-111.4	2.41	12.9		-0.75	
AUC								0.85
	6	34100828	42.9-64.1	2.26	12.1	12.1	0.67	
Oxidative Stress(300 mM H_20_2)								
G_{max}								0.91
	5	62307404	81.6-96.9	4.46	15.2	40.3	-0.88	
	6	40025376	74.9-92.6	2.12	6.7		-2.20	
	8	15684096	53.1-60.5	5.25	18.4		0.81	
t_{10}^{-1}								0.74
	2	31348124	0.0-22.4	2.3	7.6	76.3	0.11	
	4	58081284	73.1-91.0	3.78	15.3		0.82	
	6	43582592	97.2-112.7	4.21	17.3		0.63	
	7	61494964	83.2-90.6	3.4	13.6		0.58	
	8	15684096	52.2-65.6	3.59	14.3		0.90	
	10	536147	0.0-9.5	2.49	8.2		-0.90	

Table 2.1 *Continued*

	Chr[a].	Marker Peak[b]	Support Intervalc (cM)	LOD[d]	Explained Variance[e] (%)	Total Explained Variance[f] (%)	Effects[g]	Heritability[h]
	4	56773424	75.1-78.8	5.49	16.3		0.83	
	6	43582592	99.5-110.1	5.48	16.3		0.89	
	7	61494964	83.8-90.6	3.4	9.4		0.67	
	8	15684096	54.2-58.6	6.77	21.0		0.98	
AUC								0.9
	5	62100796	79.6-96.9	4.72	17.0	36.6	-0.61	
	6	39010000	27.5-110.0	2.19	3.9		0.60	
	8	15684096	52.1-63.5	4.31	15.7		0.96	

G_{max} (%), maximum germination; t_{10}^{-1}, t_{50}^{-1}, reciprocal of time to respectively 10 and 50% of viable seeds to germinate (h^{-1}); MGR, mean germination rate (reciprocal of the mean germination time; MGT^{-1}); U_{7525}^{-1}, uniformity (reciprocal of time interval between 75 and 25% viable seeds to germinate; h^{-1}); AUC, area under the germination curve (integration of fitted curve between 0 and 200 h). QTL, quantitative trait locus; RIL, recombinant inbred line; LOD, logarithm-of-odds.

[a]Chromosome number.

[b]Name (= physical position) of marker closest to the QTL peak.

[c]1-LOD support interval of QTL.

[d]LOD score that represents the significance threshold for QTL (P = 0.002) obtained by permutation tests.

[e]Percentage of variation explained by individual QTLs.

[f]Percentage of the total variance explained by genetic factors for a single trait as estimated by MapQTL.

[g]Effect of QTL calculated as mB - mA, where A and B are RILs carrying *S. lycopersicum* and *S. pimpinellifolium* alleles at the QTL position, respectively. mB and mA were estimated by MapQTL. Effects are given in percentage (G_{max}) and h^{-1} (t_{10}^{-1}, t_{50}^{-1}, MGR, U_{7525}^{-1}).

[h]Broad-sense heritability estimate for each trait, estimated as the as the proportion of phenotypic variance explained by genotype in a one-way ANOVA model; calculated as $h^2_b = (\sigma^2 g / \sigma^2 g + \sigma^2 e)$.

4, 6 and 9 which explained 33.9% of variance (Table 2.2 and 2.3, Figure 2.2). In a majority of cases the same QTLs were identified in both levels, however, there were few instances where additional QTLs were identified in one of the salt stress levels (Figure 2.2).

QTL for germination under osmotic stress conditions

QTL analysis was carried out in the case of osmotic stress for germination related traits at both low and high (-0.3 and -0.5 MPa PEG) osmotic stress conditions (Table 2.2 and 2.3). Three QTLs were identified for G_{max} under low osmotic stress on chromosomes 4, 5 and 9, whereas for high osmotic stress one QTL on chromosome 4 was identified, which explained 29.9 and 11.9% of the total variance, respectively (Table 2.2 and 2.3, Figure 2.2). For t_{10}^{-1}, two QTLs where identified for low osmotic stress on chromosomes 2 and 4, which explained 27.2% of total variance, whereas at high osmotic stress five QTLs were identified, one each on chromosomes 2, 4, 6, 9 and 12 with a total explained variance of 62.3% (Table 2.2 and 2.3, Figure 2.2). The U_{7525}^{-1} was not calculated as the final germination percentage was too low to calculate meaningful values for the corresponding fraction, as previously described. Four QTLs were identified for AUC in case of low osmotic stress on chromosome 4, 6, 9 and 12 and four QTLs were detected at high osmotic stress conditions, one each on chromosomes 2, 4, 6 and 12 (Table 2.2 and 2.3, Figure 2.2), which accounted for 48.1 and 44.1% of the total explained variance, respectively. Similar as described for salt, in a majority of cases the same QTLs were identified in both levels, however, there were few instances where additional QTLs were identified in one of

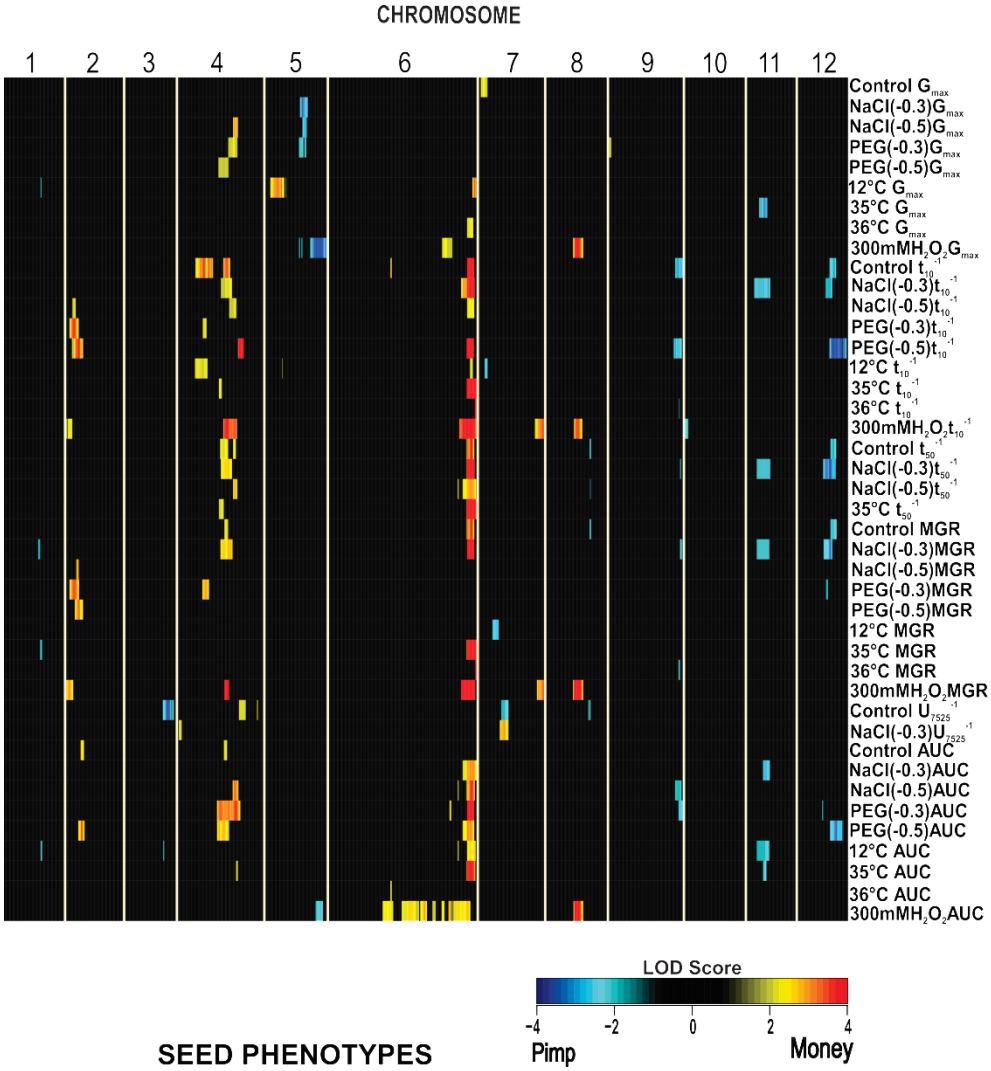


Figure 2.2. Genomic locations of quantitative trait locus (QTL) identified for seed quality traits. Tomato chromosomes are identified by numbers (1–12), with centimorgans ascending from the left to right; chromosomes are separated by white lines. Control indicates germination phenotypes under optimal condition. Colored cells indicate QTL significant at P = 0.002 in multiple QTL mapping models [1-logarithm-of-odds (LOD)]. The LOD color scale is indicated, showing blue and light blue when the *Solanum*

pimpinellifolium (Pimp) allele, and yellow and red when the *Solanum lycopersicum* (Money) allele, at that marker results in an elevated level of seed quality phenotype. QTL positions, LOD scores, effects and h_b values are provided in Table 2.2. AUC, area under the germination curve; MGR, mean germination rate.

the osmotic stress levels (Figure 2.2).

QTL for germination under temperature stress conditions

Cold stress. Three QTLs were found for G_{max} at 12 °C on chromosomes 1, 5 and 6, which accounted for 32.4% of the total explained variance (Table 2.2 and 2.3, Figure 2.2). For t_{10}^{-1}, four QTL were found on chromosome 4, 5, 6 and 7 with 42.1% of total explained variance (Table 2.2 and 2.3, Figure 2.2), whereas U_{7525}^{-1} was not obtained as the final germination percentage was too low to calculate meaningful values. Four QTLs were found for AUC at 12 °C on chromosomes 1, 3, 6, and 11 with 37.0% of total explained variance.

High temperature. One QTL each on chromosomes 11 and 6 was found for G_{max}, at 35 °C and 36 °C, which explained 14.7% and 12.7% of the variance, respectively (Table 2.2 and 2.3, Figure 2.2). One QTL each on chromosomes 4 and 6 for t_{10}^{-1} was identified at 35 °C whereas two QTLs on chromosomes 6 and 9 at 36 °C were found, which explained 28.7% and 42.5% of the total variance, respectively (Table 2.2 and 2.3, Figure 2.2). U_{7525}^{-1} was not calculated as the majority of RILs did not reach a final germination percentage above 75 %. Three QTLs were found one each on the chromosomes 4, 6 and 11 for AUC at 35 °C and 1 QTL on chromosomes 6 for AUC at 36 °C, which explained 30.7 and 12.1% of the

Table 2.3. Summary of QTL of seed quality traits in *S. lycopersicum* / *S. pimpinellifolium* RIL population

Treatments	Traits[a]	QTL (nr)[b]	Range of Explained Variance (%)[c]	Total Explained Variance(%)[d]
Control				
	G_{max}	1	11.9	11.9
	t_{10}^{-1}	5	11.8-20.6	59.3
	t_{50}^{-1}	4	7.6-13.1	36.8
	MGR	4	7.2-12.9	36.6
	U_{7525}^{-1}	4	7.2-12.8	42.9
	AUC	2	10.3-12.3	22.6
Salt Stress I (-0.3 MPa NaCl)				
	G_{max}	1	15.7	15.7
	t_{10}^{-1}	4	8.1-11.2	39.3
	t_{50}^{-1}	5	6.4-13.3	52.1
	MGR	6	5.6-17.4	57.7
	U_{7525}^{-1}	2	9.0-13.2	22.2
	AUC	2	12.4-13.7	26.1
Salt Stress II (-0.5 MPa NaCl)				
	G_{max}	2	12.7-14.4	27.1
	t_{10}^{-1}	3	7.2-7.8	22.7
	t_{50}^{-1}	3	8.0-11.6	31
	MGR	1	15.7	15.7
	AUC	3	8.0-13.4	33.9
Osmotic Stress I (-0.3 MPa PEG)				
	G_{max}	3	8.6-11.2	29.9
	t_{10}^{-1}	2	10.4-16.8	27.2
	MGR	3	8.2-14.8	34.8
	AUC	4	7.0-18.7	48.1
Osmotic Stress II (-0.5 MPa PEG)				
	G_{max}	1	13.5-13.5	11.9
	t_{10}^{-1}	5	7.9-16.6	62.3
	MGR	1	13.5-13.5	13.5

Table 2.3 *Continued*

Treatments	Traits[a]	QTL (nr)[b]	Range of Explained Variance (%)[c]	Total Explained Variance(%)[d]
	AUC	4	9.8-12.3	44.1
Cold Stress (12 °C)				
	G_{max}	3	8.8-14.0	32.4
	t_{10}^{-1}	4	9.1-11.9	42.1
	MGR	1	12.6	12.6
	AUC	4	7.5-10.7	37
Temperature Stress I (35 °C)				
	G_{max}	1	14.7	14.7
	t_{10}^{-1}	2	10.0-23.0	33
	t_{50}^{-1}	2	9.2-19.5	28.7
	MGR	2	11.6-23.4	35
	AUC	3	10.0	30.7
Temperature Stress II (36 °C)				
	G_{max}	1	12.7	12.7
	t_{10}^{-1}	2	10.9-31.6	42.5
	MGR	2	12.9-28.8	41.7
	AUC	1	11.6	12.1
Oxidative Stress (300mM H_2O_2)				
	G_{max}	3	6.7-18.4	40.3
	t_{10}^{-1}	6	7.6-17.3	76.3
	MGR	5	6.5-16.3	69.5
	AUC	3	3.9-17.0	36.6

[a]G_{max} (%), maximum germination; t_{10}^{-1}, t_{50}^{-1} reciprocal of time to respectively 10 and 50 % of viable seeds to germinate (h^{-1}); MGR, mean germination rate (reciprocal of the mean germination time; MGT^{-1}); U_{7525}^{-1}, uniformity (reciprocal of time interval between 75 and 25% viable seeds to

germinate (h^{-1}); and AUC, area under the germination curve (integration of fitted curve between 0 and 200 hours).

[b] Number of QTLs detected

[c] Range of explained variance for QTLs

[d] Total explained variance for each trait

variation, respectively (Table 2.2 and 2.3, Figure 2.2).

QTL for germination under oxidative stress conditions

Three QTLs were identified for G_{max} on chromosomes 5, 6 and 8 for oxidative stress, which explained 40.3% of the total variance (Tables 2.2 and2.3, Figure2.2). QTL analysis revealed six QTLs for t_{10}^{-1} on chromosomes 2, 4, 6, 7, 8 and 10 with 76.3% of the total explained variation (Tables 2.2 and 2.3, Figure 2.2). No estimate for U_{7525}^{-1} was obtained as the final germination percentage was too low to calculate meaningful values. For AUC, three QTLs were found on chromosomes 5, 6 and 8 accounting for 36.6% of the total explained variance (Tables 2.2 and 2.3, Figure2.2).

Shared QTLs among seed phenotypes

Permutation tests conducted onto all -1LOD QTL intervals allowed to compare and estimate the level of overlapping QTLs between phenotypic traits where occurrences of overlapping QTLs between different seed quality traits considered highly significant with 1-P-value of 0.99 or 1.0. Seven QTL clusters positioned onto chromosomes 1, 2, 4, 6, 8, 9 and 12 were identified as affecting different seed germination traits with an overlapping proportion ranging from 62.5 to 100% at -1LOD (Figure 2.2). QTLs positioned onto chromosomes 1, 2, 4, 6, 9 and 12 also revealed at -

1LOD a significant overlap (from 91.6 to 100%) between QTL clusters for rate of germination parameters (t_{10}^{-1}, t_{50}^{-1}, MGR). QTLs detected for G_{max}, t_{10}^{-1}, t_{50}^{-1} and MGR co-located significantly onto three chromosomes: chromosomes 6, 9, and 12 (Figure 2.2). The overlapping range between QTLs affecting simultaneously t_{10}^{-1}, t_{50}^{-1} and MGR varied from 90.0 to 100% (Figure 2.2). QTLs involving G_{max} and AUC traits co-located together onto the chromosomes 4, 6, and 11, whereas AUC and t_{10}^{-1}, t_{50}^{-1} and MGR QTLs were significantly overlapping (from 79.4 to 100%) onto chromosomes 3, 4, 6, 8, 9, 11, and 12 (Figure 2.2).

To investigate associations among characteristics at the phenotypic level, a correlation matrix was generated by performing Pearson correlation analysis for all pairs of measured traits across the whole population. This analysis used average values calculated from all raw determinations for a given trait/RIL pair. Pearson correlation coefficients (Rp) and accompanying false discovery rate (FDR)–corrected P values (P_{BH}; Benjamini and Yekutieli, 2001) are provided in Supporting Information Table S2.1. Using the Pearson correlation coefficient to calculate relationships among seed quality phenotypes concerned, a number of low to high significant correlations were observed for seed phenotypes under different germination conditions(Figure 2.3 and Supporting Information Figure S2.1, Supporting Information Table S2.1).For instance, G_{max} in almost all germination conditions was slightly to highly correlated with t_{10}^{-1}, t_{50}^{-1} and U_{7525}^{-1} (R_p = 0.49 to 0.76; P_{BH} = 0.00). In case of AUC, significant correlations were also observed between these traits (up to R_p = 0.87; P_{BH} = 0.00). Significant positive correlations were also observed between the G_{max} and AUC under different germination conditions. Furthermore, there was a strong correlation between the t_{10}^{-1}, t_{50}^{-1}, MGR and U_{7525}^{-1}(P_{BH}< 0.0001) (Figure 2.3 and Supporting Information Figure S2.1, Supporting Information Table

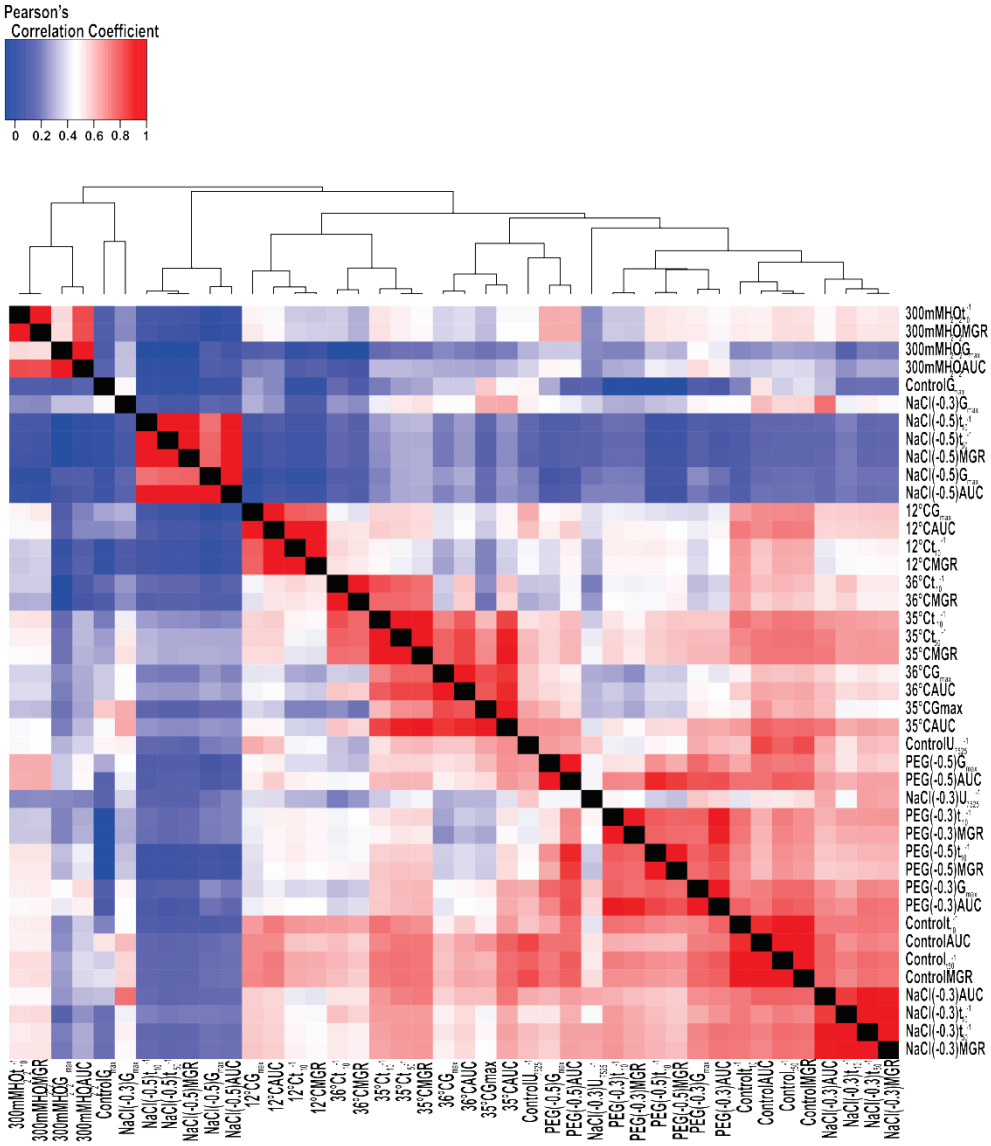


Figure 2.3. Heatmap of correlations between seed quality phenotypes. Each square represents the Pearson correlation coefficient between the seed phenotypes of the column with that of the row. Seed phenotype order is determined as in hierarchical clustering using the distance function 1-correlation. The dissimilarity index is employed for cluster analysis to arrange different seed phenotypes according to their similarity (Legendre &

Legendre 1998). Self-self correlations are identified in black. Individual correlation coefficients can be found in Supporting Information Table S2.1. Supporting Information Figure S2.1 displays the correlation heatmap organized in logical order for calculated seed traits, for example, G_{max}, t_{10}^{-1}, t_{50}^{-1}, MGR, U_{7525}^{-1} and AUC. AUC, area under the germination curve; MGR, mean germination rate.

S2.1). This is most obvious between t_{10}^{-1} and t_{50}^{-1}. Examples of t_{10}^{-1}-t_{50}^{-1} correlations include control- t_{10}^{-1} and t_{50}^{-1} (R_p = +0.95; P_{BH} = 0.00), salt- (-0.3MPa, -0.5MPa), $t_{10}^{-1}$and t_{50}^{-1} (R_p = +0.95; P = 0.00; R_p = +0.97; P_{BH} = 0.00 respectively), and between t_{10}^{-1} and t_{50}^{-1} at high-temperature stress-(35°C) (R_p = +0.97; P_{BH} = 0.00).The trend was similar while comparing MGR with G_{max}, t_{10}^{-1}, t_{50}^{-1}, and AUC; a number of low to high significant correlations were observed for seed phenotypes under different germination conditions (Figure 2.3 and Supporting Information Figure S2.1, Supporting Information Table S2.1).

Epistasis

The results of genome-wide epistasis analysis for each of the seed quality phenotypes are presented in Table 2.4. These analyses tested all pairwise combinations of the markers closest to each target QTL. The analysis of this interaction among seed quality QTL revealed several instances where epistatic interactions among QTLs may obscure relationships between loci and phenotypes. These epistatic interactions contribute to phenotypic variability, but hinder detection and affect estimation of QTLs examined singly. A survey of epistasis with the R\qtl module detected reasonable instances of epistasis in our experiments, whereby only pairwise interactions involving two loci were tested. This

analysis revealed novel loci on several chromosomes interacting to influence seed quality traits.

The analysis revealed a locus on chromosomes 4 and 5 interacting to influence U_{7525}^{-1} under control conditions (Table 2.4, Figure 2.4). Similarly, for salt- (-0.3 MPa), strong evidence of interaction was observed for U_{7525}^{-1} on chromosomes 4 and 7 (LOD_{int}= 5.00). This was the highest level of statistical significance obtained in our epistasis screen. A two-way interaction was also revealed for t_{10}^{-1} on chromosomes 2 and 4 under salt-stress conditions (-0.5 MPa), whereas a locus on chromosome 2 also interacts with a locus on chromosome 4 under osmotic-stress condition (-0.3MPa PEG) for the same parameter (Table 2.4, Figure 2.4). An epistatic interaction was also observed for AUC under cold stress (12 °C) between QTL on chromosome 3 and 11 (Table 2.4, Figure 2.4).

Table 2.4. Interaction LOD scores for phenotypes significant at the genome-wide level (P < 0.05)

Phenotype	ChrA	Position (cM)	ChrB	Position (cM)	Lod.full[a]	Lod.fv1[b]	Lod.int[c]
Control U_{7525}^{-1}	4	85	5	15	7.62	6.0	4.56
Salt I (-0.3) U_{7525}^{-1}	4	10	7	52	10.41	7.43	5.00
Salt II (-0.5) t_{10}^{-1}	2	25	4	65	9.48	6.62	4.00
Osmotic I (-0.3) t_{10}^{-1}	2	22	4	25	11.97	6.62	4.55
Cold Stress (12 °C) AUC	3	55	11	15	8.98	4.7	4.91

Two-way epistatic interactions for *Solanum lycopersicum/Solanum pimpinellifolium* recombinant inbred line population across all 12 chromosomes. AUC, area under the germination curve; LOD, logarithm-of-odds.

[a]Lod.full is the LOD score of the full model with two loci and their interaction compared with the null model with no quantitative trait locus

(QTL).

[b]Lod.fv1 is the LOD score of the full model compared with the best single QTL model with one locus on either chromosome A or B.

[c]Lod.int is the LOD score of the interaction term which is found by comparing the full model with an interaction term to the two QTL models with no interaction term.

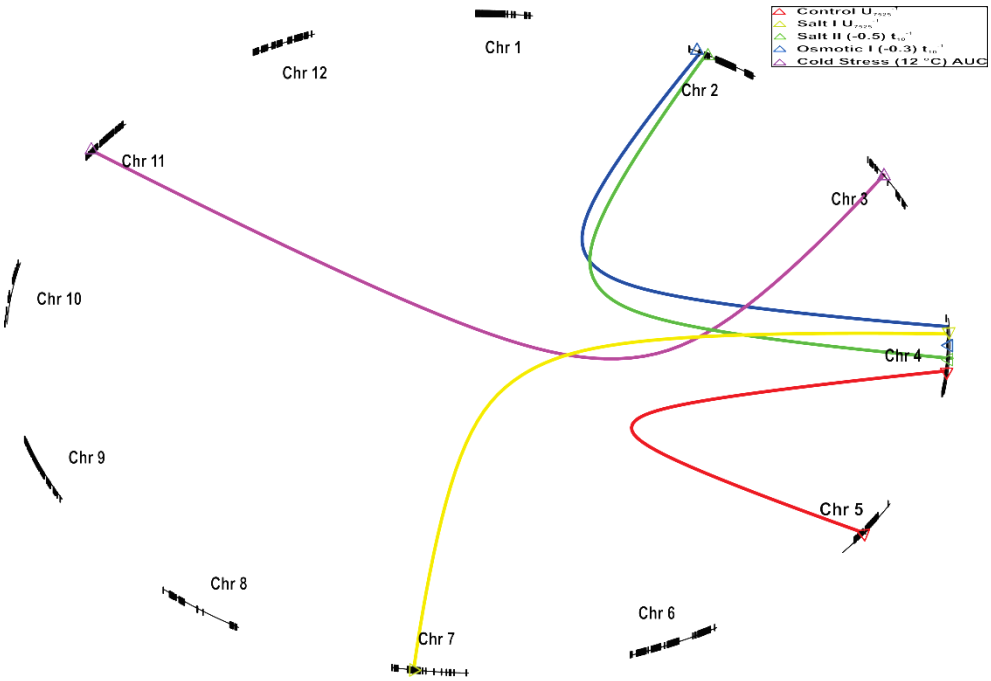


Figure 2.4. Epistatic interaction network. Graphical visualization of the epistatic interactions found between different loci controlling seed quality phenotypes in *Solanum lycopersicum* and *Solanum pimpinellifolium* recombinant inbred line population. The 12 chromosomes are represented as different circle segments, and their sizes are proportional to the corresponding genetic sizes measured in centimorgan (cM) units. The color of the lines indicates the trait for which the epistatic interaction was observed (Arends *et al.* 2010). AUC, area under the germination curve.

Discussion

This study makes clear that the genetic control of seed quality is complex. We have detected numerous QTLs with moderate to large phenotypic effects that influence tomato seed quality attributes consistently across all studied traits. Contributions to seed quality from both tomato parental genotypes produced transgressive segregation for some traits. We also found significant evidence for pairwise epistatic interactions. Differences in QTL detection among phenotypic traits added new dimensions to the complexity of seed quality. The recognition and assessment of sources of variation of seed quality is essential for developing a realistic understanding of how tomato seed phenotypes interact across different conditions, with the ultimate goal of obtaining durable seed quality in tomato crop plants.

The *S. lycopersicum* × *S. pimpinellifolium* RIL population and QTL locations

The power of detecting QTLs depends on several factors, including heritability (h^2) of the trait, gene action, the type of mapping population, the number and individual effects of QTLs, marker coverage and the distance between marker loci and QTL(s) affecting the trait (Mackay, 2001; Foolad et al., 2003; Mackay et al., 2009). The overall heritability of traits (i.e. heritability in the broad sense) strongly affects the quality of QTL analysis, including the number of QTLs detected and the accuracy of their map positions and effect estimates (Alonso-Blanco and Koornneef, 2000). However, heritability in the broad sense can be controlled by several factors, which are experimentally manipulable when scoring the traits (Kobayashi and Koyama, 2002). We have utilized homogenous and strictly controlled plant growing and seed phenotype testing conditions and this

has contributed to increasing the broad sense heritability of the seed quality traits in both control- and stressed conditions (h_b^2>0.53-0.94; Table 2.2).

Interpretation of seed germination traits

Several methods and mathematical expressions to measure the germination process have been proposed over the past two decades (Hilhorst and Karssen, 1988; Bradford, 1990; Bewley and Black, 1994). One of the most significant current discussions in seed science concerns the measurement of time, rate, homogeneity, and synchrony of germination, as they can provide information about the dynamics of the seed germination process. These characteristics are important for physiologists and seed technologists as it is at the heart of their understanding of germination potential of seedlots. This study is an effort of indexing different aspects of cumulative germination in order to quantify the different seed quality traits under different germination conditions. The final germination of seeds is one of the qualitative attributes of the germination process; it portrays the overall germination potential of crop species based on a binary answer: germinated or non-germinated. There is consensus as to the meaning, methods and calculation of germinability in time or at the end of the observations (Ranal and Santana, 2006). Although final germination is an important factor for estimating the expected seedling yield of a seedlot, it can be partly independent of other germination characteristics like rate of germination. The germination characteristics of a seedlot are determined by the species, genetic diversity as well as germination conditions and seed pre-treatments. In fact, it has been shown that germination parameters are under strong genetic control (El-Kassaby, 1991) and therefore, analyzing different aspects of cumulative

germination curves, like the onset of germination and germination rate as important phenotypic attributes of a seedlot is of unprecedented importance with respect to the consequences of genetic diversity present in the *S. lycopersicum* x *S. pimpinellifolium* RIL population. However, it has been emphasized that onset of germination and germination rate (t_{10}^{-1}, t_{50}^{-1} respectively) are useful for comparisons only when samples have a sufficient level of final germination (Goodchild and Walker, 1971), and to address this issue, we only measured these parameters for those traits that show at least 10 and 50% germination respectively in more than 80% of the RILs. There is a large volume of published studies describing genetic characterization of onset and rate of seed germination (t_{10}^{-1}, t_{50}^{-1}, MGR) and exploitation of the natural variation using different mapping populations e.g. RILs, ILs etc. for germination rate phenotypes (Quesada et al., 2002; Foolad et al., 2003; Foolad et al., 2003; Clerkx et al., 2004; Langridge et al., 2006; Foolad et al., 2007; Landjeva et al., 2010). In this study we performed QTL analysis with all these different germination parameters and we found genomic regions where QTLs for different rate measurements were mapped to the same approximate location, indicating that common factors are associated with the rate measurements to different germination conditions. Strong correlations were also evident among the different rate measurements, and Pearson's correlation analysis among all rate estimates indicated high correlations among t_{10}^{-1}, t_{50}^{-1}, and MGR ($P < 0.0001$).

Despite the agronomic importance of the rate and uniformity of germination, these traits have not been specifically targeted by breeders. Longer germination times for tomato seeds have been associated with a greater likelihood of producing an abnormal seedling. In terms of seed vigor, the rate and uniformity of germination is a sensitive indicator of a

high-quality seed, and these attributes deteriorate more quickly than final germination and are therefore a key component to seed quality. To simplify quantification of germination responses, both the rate and percentage of germination were incorporated into AUC. Thus, simultaneous germination responses can be interpreted by the AUC as increases in germination rate and final germination percentage, as well as an earlier onset and uniformity of germination. Seedlots that germinate rapidly and fully will have high AUC values, while those that germinate slowly and lowly will have low values. The analysis of germination can be enriched if, in addition to the final germination, t_{10}^{-1}, t_{50}^{-1}, MGR, U_{7525}^{-1} and AUC values are communicated, because they measure different aspects of the germination process. t_{10}^{-1} which is predominantly a measure of the onset of germination (lag time) whereas t_{50}^{-1} and MGR are measures for the germination rate, U_{7525}^{-1} for uniformity and AUC as the combinatorial parameter. This study demonstrates the usefulness of these germination parameters for describing the extremes of pattern differences of seed germination and all these germination measurements can be applied to evaluate seed germination.

QTL overlapping among seed quality phenotypes

Because seed quality is attributable to an overall tolerance to various seed stresses we expected, and found evidence for, the co-location of QTLs for control and all stress conditions. A number of significant occurrences of overlapping QTLs among G_{max}, t_{10}^{-1}, t_{50}^{-1}, MGR, U_{7525}^{-1} and AUC were observed among most of the detected QTL positions across different germination conditions. For instance, on chromosomes 4, 6 and 11 the confidence intervals of G_{max} and AUC QTLs overlapped with those detected for t_{10}^{-1}, t_{50}^{-1}, and MGR across different stress conditions (Figure

2.2). Another instance of significant co-locations of QTLs was identified for these seed quality traits on linkage groups 1, 2, 9 and 12 (Figure 2.2). Such co-locations indicate that the shared QTL clusters may bear pleiotropic effects. The co-locations of QTLs identified for seed quality traits in the present study indicated a variable number of overlapping QTL clusters among them. The co-location of roughly two-thirds of the QTLs affecting the t_{10}^{-1}, t_{50}^{-1} and MGR across different stresses highlights the positive relationship between seed quality phenotypes and different stress types. The present results indeed corroborate previous QTL mapping studies of germination under salt-, drought- and cold- stresses in tomato where 71% of the detected QTLs affected germination under two stresses or more (Foolad et al., 2007). Although, QTLs for the seed quality parameters (G_{max}, t_{10}^{-1}, t_{50}^{-1}, MGR, U_{7525}^{-1}, and AUC) in each germination condition often co-located as partly may be explained by the fact that they all are descriptors for the same germination curves, interestingly however in several instances, germination parameters mapped to unique regions e.g. QTLs for G_{max} on chromosome 5 at 12 °C, t_{10}^{-1} under oxidative stress on chromosome 10 and QTL for U_{7525}^{-1} under control condition on chromosome 3 (Figure 2.2). Furthermore, inspection of the QTLs affecting individual parameters across different chromosomes also revealed striking significant hot spots for one parameter but not for other. Examples include on chromosome 5 we had QTLs for G_{max}, but not for t_{10}^{-1}, t_{50}^{-1} or MGR, whereas on chromosome 7 we had co-location for t_{50}^{-1} and MGR, but no revelation of any QTL for G_{max}. Furthermore, overlapping QTLs were found on chromosome 9 for t_{10}^{-1}, t_{50}^{-1} and AUC, but not for other measured traits. Similarly on chromosome 12 we had QTL overlaps for t_{10}^{-1}, t_{50}^{-1}, MGR and AUC traits but not for G_{max} (Figure 2.2). Apparently there are specific loci that affect some germination characteristics and not others. It is also

interesting to note that besides QTLs at the same loci for all salt and osmotic levels, in some instances additional QTLs under certain concentration were revealed (chromosomes 2, 4, 5, 6, 9, 11 and 12). As an example G_{max} QTL on chromosome 5 was detected in both salt stress levels whereas a QTL on chromosome 4 was only detected at -0.5 MPa salt. The magnitude of different stresses is variable in soil and stress tolerance to environmental stresses depends on the stage, length and severity of the stress (Bray, 2002). These results indicate that seeds respond to one or more stresses through physiological mechanisms depending on the nature and magnitude of the stress (Capiati et al., 2006). Similarly, while comparing QTLs for salt and osmotic stress conditions we found QTLs co-locating for some seed germination parameters for both salt and osmotic stress, but we were also able to identify novel loci (Figure 2.2). These findings further support the idea of that the regulation of germination under salt and osmotic stresses involves the action of common as well as independent loci, revealing the existence of loci specifically associated with the toxic component of salt and not just its osmotic effect (Vallejo et al., 2010). Furthermore, identification of QTLs for non-stress condition indicates the genetic relationships between germination phenotypes under stress- and non-stress conditions and it has been suggested that germination of tomato is genetically controlled and hence can be increased by selection (Foolad et al., 1999). QTLs corresponding to different seed parameters in our study have shown overlaps, and correlations among germination-derived parameters were also high. Thus, establishing the correspondence between QTL co-locations and correlations between phenotypic characters appears possible. Considering together the traits studied herein, significant correlations were observed: up to 0.76 between G_{max} and AUC, and up to 0.95 between G_{max} and t_{10}^{-1}, t_{50}^{-1}, MGR and U_{7525}^{-}

[1]and likewise up to 0.87 between AUC and aforementioned parameters. The QTL analysis indicated the presence of genetic relationships between germination under different conditions. These observations suggest that the QTLs detected for G_{max}, t_{10}^{-1}, t_{50}^{-1}, MGR, U_{7525}^{-1} and AUC in tomato seed are overlapping on the same linkage groups and could be related to significant correlations among these traits. Previous quantitative trait genetic studies have reported similar co-locations (Foolad and Chen, 1999; Clerkx et al., 2004) and suggest that trait correlations may be attributable to either pleiotropic effects of single genes or to tight linkage of several genes that individually influence specific traits (Pelgas et al., 2011). It should not be too difficult to disentangle these two effects in the near future.

Physiological mechanism of seed quality phenotypes under different conditions

Productive and sustainable crop growth necessitates growing plants in sub-optimal environments with less input of precious resources. This study was intended to make a step forward towards better understanding and rapid improvement of abiotic stress tolerance in tomato, and to link physiological and underlying molecular mechanisms of seed quality. Excessive salt lowers the rate of, or completely inhibits, seed germination (Foolad et al., 2003; Foolad et al., 2007). This may be accomplished by lowering the osmotic potential of the germination medium, but a saline germination medium could also lower the rate of seed germination by specific salt stress. However, accumulating evidence suggests that the low water potential of the external medium, rather than ion toxicity effects, is the major limiting factor to germination under salt stress in different crop species, including tomato (Ni and Bradford, 1992; Bradford, 1995; Foolad

et al., 2007). Another possible explanation for some of our results may be the release of reactive oxygen species (ROS) in all of these stress types (Clerkx et al., 2004; Wahid et al., 2007; Collins et al., 2008). Saline conditions are known to generate ROS (Zhu, 2002). Prior studies have noted that lowered rates of seed germination under drought stress are due to reduced osmotic potential of the germination medium (Bradford, 1995; Hilhorst and Downie, 1996) similar to that under salt stress. Therefore, it is expected that seeds that germinate rapidly under salt stress would also germinate rapidly under osmotic stress, and vice versa. This is partly in agreement with the findings of the present study. It is conceivable that similar or identical genes (and physiological mechanisms) control the seed germination process of tomato under salt and drought stress. This is evident from the correlation between salt and PEG treatments (Figure 2.3). There is hardly any information whether genetic and physiological processes that maintain rapid seed germination under salt and/or drought stress are also responsible for rapid seed germination under cold stress. However, low temperature (cold stress) may affect the water status of the cell and, thus, could delay seed germination by causing osmotic stress (Liptay and Schopfer, 1983). In the present study, however, the finding that most of QTLs for seed quality traits under cold stress co-localized with QTLs for germination under salt and/or osmotic stress suggests that the same genes (or physiological mechanisms) may contribute to rapid seed germination under these three conditions. This suggestion is consistent with the finding that selection for rapid seed germination under salt or drought stress resulted in progeny with improved germination under cold-stress, and vice versa (Foolad and Lin, 2000).

In the present study, QTLs were identified affecting germination phenotypes under non-stress- (control) and stress conditions (Figure 2.2).

The QTLs located on chromosomes 4, 6 and 11 affected germination under three or more conditions. Correlation analysis indicated highly significant correlations between the various germination traits at all treatment levels and this suggests that for response time traits like germination, the earlier traits may be good predictors of crop performance. Genes related to reserve mobilization and endosperm weakening are likely to be involved and these could conceivably affect the rate of germination as metabolic processes and reserves utilized early during germination are different from those required later during the process, but before its completion (Fait et al., 2006; Bethke et al., 2007; Hayashi et al., 2008), and indeed, presence of QTLs for different germination phenotypes, in particular t_{10}^{-1}, t_{50}^{-1} and MGR on different chromosomes of the tomato genome possibly corresponds to metabolic or physiological processes that are themselves occurring during different stages of the germination process. A number of QTLs associated with time to 50 % of germination (t_{50}) were mapped in tomato (Foolad et al. 1999, 2003), Arabidopsis (Quesada et al., 2002) and 1 QTL was also mapped in sunflower (Al-Chaarani et al., 2005).

This study clearly illustrates the complexity underlying the genetic basis for seed germination. Identifying QTLs associated with the different parameters of seed germination facilitates elucidation of molecular mechanisms controlling seed germination. As suggested by transgressive trait distributions within the RILs, both parental genotypes *S. lycopersicum* cv. Moneymakerand *S. pimpinellifolium* contributed to increased trait means for different germination parameters under control (non-stress) and stressed conditions. This phenomenon has frequently been described for other traits in many crops (Devicente and Tanksley, 1993; Foolad, 1996), including tomato. The presence of favorable alleles in both parents suggests a strong likelihood for recovering transgressive variants among

segregating progeny (Devicente and Tanksley, 1993). Given the result that alleles serving to enhance the ability to complete germination under environmental stress are present in both cultivars, improvement of germination traits must be conducted at an individual QTL level (Hayashi et al., 2008).

Detection of QTLs generic to germination traits under control and stressed conditions suggests the presence of genetic relationships between the ability to germinate rapidly under different conditions and the prediction that selection and improvement of seed germination under one condition would lead to progeny with improved germination under other conditions. There was evidence of greater germination variances in the current study under stress conditions, which is partly due to slower germination and, thus, longer time intervals between germination events. Under stress conditions, germination variances increased in the RIL population, and broad sense heritabilities were larger under stress than non-stress conditions, suggesting the contribution of some genetic factors to the larger variance under the stress treatment. Greater genetic variance in stress environments is rather uncommon, but is one of the more favorable situations for plant breeders (Rosielle and Hamblin, 1981). Furthermore, seed germination under different stress conditions was genetically controlled with additivity being the major genetic component. Significantly large genetic correlations between germination responses at different stress levels indicate that similar or identical genes contributed to the germination response under different stress conditions. Thus, selection for rapid germination at one stress level would result in progeny with improved germination at diverse stress levels. Nonetheless, the co-location of QTLs for different seed germination traits supports the genetic dissection of seed quality in order to facilitate a more strategic approach to breed for

better seed quality in tomato. Those regions identified across different germination environments are candidates that can be used in marker assisted selection (MAS) or gene cloning, especially those with moderate to high broad sense heritabilities (Dudley, 1993; Tanksley, 1993). However, isolation, characterization, and comparison of functional genes, which facilitate rapid seed germination under the various conditions, are necessary in order to determine the exact genetic relationships among these traits.

Identification of epistasis

We have performed a genome-wide epistasis screen in the *S. lycopersicum* x *S. pimpinellifolium* cross for seed quality phenotypes and obtained evidence for multiple significant QTL pairs. The identification of significant epistasis controlling seed quality phenotypes both benefits and complicates this analysis. Epistasis may identify genes that function together in distinct genetic networks, potentially providing a valuable insight into function. Our identification of higher-order epistatic networks that control quantitative seed quality phenotypes in *S. lycopersicum* suggests that these QTLs may be caused by polymorphism in genes that function in a coordinated network. These findings exemplify an advantage of interaction analyses in plant models for complex phenotypes such as seed quality, since by the use of R/qtl analysis we had more than sufficient statistical power to detect 2 way epistatic interactions, implicating genomic regions that would otherwise likely have been passed over (Buescher et al., 2010).

Identification of epistatic pairs of loci contribution to seed quality variation in tomato represents a step forward in the delineation of the genetic architecture of these phenotypes in tomato and provides a powerful

approach to identify novel gene candidates and chromosomal regions for further pursuit in seed quality studies. Our results, however, also illustrate the degree of complexity of the genetic architecture of these phenotypes. Strong epistasis in the genetic network controlling germination under salt stress was revealed in an Arabidopsis Sha x Col RIL population (Galpaz and Reymond, 2010). Validation of this epistatic network hypothesis will require cloning of the full complement of interacting QTLs. Accounting for these seed quality QTL interactions is not only essential for developing strategies to clone seed quality QTLs, but may also allow the useful inclusion of metabolomics and transcriptomics data in the formulation of hypotheses regarding mechanisms of seed quality of the tomato.

In conclusion, this study has identified numerous QTLs contributing to variation in seed quality trait interactions between the tomato accessions *S. lycopersicum* and *S. pimpinellifolium*. The QTL approach appears to be valuable not only in elucidating the genetics, but also the physiological background of the seed quality phenotypes. Both stress-specific and non-specific QTLs control the germination process under different conditions in the tomato. This approach offers a way in which simultaneous improvement of these traits and progress toward identifying the underlying genetic mechanisms may be realized. Genome-scale prediction of a large-effect DNA sequence and transcript accumulation polymorphisms differentiating *S. lycopersicum* and *S. pimpinellifolium* permit an informed approach to selection and investigation of gene candidates in identified QTL regions (Joosen et al., 2009). The present study is a significant effort in this direction. Robust QTL mapping with SNP-based linkage maps resulted in a much-improved estimation of the genetic architecture of a tomato genome in terms of the magnitude of QTL effects, QTL-environment interactions, and putative pleiotropy. Identification of causal polymorphisms

for QTLs influencing a majority of *S. lycopersicum* and *S. pimpinellifolium* phenotypes will provide potential breeding targets for enhanced seed quality in tomato. Furthermore, fine mapping, validation and further investigation of seed quality-specific QTL swill provide valuable insight into pleiotropic variation as suggested by the co-location of the QTLs.

Acknowledgments
This work was supported by the Technology Foundation STW (R.K., L.W., W.L.) and by the Higher Education Commission, Pakistan (N.K.).

Supporting Information
Supporting information can be downloaded from either the online version of this article (Kazmi et al. 2012) or fromwww.wageningenseedlab.nl/thesis/rhkazmi/chapter2

Table S2.1: Self-self correlations are identified in black. Individual correlation coefficients can be found in Table S2.1.

Figure S2.1 displays the correlation heatmap organised in logical order of calculated seed traits e.g. G_{max}, t_{10}^{-1}, t_{50}^{-1}, MGR, U_{7525}^{-1}, AUC.

References

Al-Chaarani GR, Gentzbittel L, Wedzony M, Sarrafi A (2005) Identification of QTLs for germination and seedling development in sunflower (Helianthus annuus L.). Plant Science **169:** 221-227

Alonso-Blanco C, Aarts MGM, Bentsink L, Keurentjes JJB, Reymond M, Vreugdenhil D, Koornneef M (2009) What Has Natural Variation Taught Us about Plant Development, Physiology, and Adaptation? Plant Cell **21:** 1877-1896

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

Asghari A (2007) QTL analysis for cold resistance-related traits in Brassica napus using RAPD markers. International journal of food, agriculture and environment **5:** 188-192

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. Annals of statistics**:** 1165-1188

Bentsink L, Hanson J, Hanhart CJ, Blankestijn-de Vries H, Coltrane C, Keizer P, El-Lithy M, Alonso-Blanco C, de Andres MT, Reymond M, van Eeuwijk F, Smeekens S, Koornneef M (2010) Natural variation for seed dormancy in Arabidopsis is regulated by additive genetic and molecular pathways. Proceedings of the National Academy of Sciences of the United States of America **107:** 4264-4269

Bethke PC, Libourel IGL, Aoyama N, Chung YY, Still DW, Jones RL (2007) The Arabidopsis aleurone layer responds to nitric oxide, gibberellin, and abscisic acid and is sufficient and necessary for seed dormancy. Plant Physiology **143:** 1173-1188

Bewley JD (1997) Seed germination and dormancy. Plant Cell **9:** 1055-1066

Bewley JD, Black M (1994) Seeds: physiology of development and germination. Springer, New York

Borlaug NE, Dowswell CR (2003) Feeding a world of ten billion people: a 21st century challenge. . *In* R Tuberosa, RL Phillips, M Gale, eds, Proceedings of the International Congress in the Wake of the

Double Helix: From the Green Revolution to the Gene Revolution. Avenue Media, Bologna, pp 3-23

Bradford KJ (1990) A water relations analysis of seed-germination rates. Plant Physiology **94:** 840-849

Bradford KJ (1995) Water relations in seed germination. *In* J Kigel, G Galili, eds, Seed development and germination. Marcel Dekker Inc., New York, pp 351-396

Bray EA (2002) Abscisic acid regulation of gene expression during water-deficit stress in the era of the Arabidopsis genome. Plant Cell and Environment **25:** 153-161

Broman KW, Wu H, Sen S, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. Bioinformatics **19:** 889-890

Buescher E, Achberger T, Amusan I, Giannini A, Ochsenfeld C, Rus A, Lahner B, Hoekenga O, Yakubova E, Harper JF, Guerinot ML, Zhang M, Salt DE, Baxter IR (2010) Natural Genetic Variation in Selected Populations of Arabidopsis thaliana Is Associated with Ionomic Differences. PLoS ONE **5:** e11081

Campos H, Cooper A, Habben JE, Edmeades GO, Schussler JR (2004) Improving drought tolerance in maize: a view from industry. Field Crops Research **90:** 19-34

Capiati DA, Pais SM, Tellez-Inon MT (2006) Wounding increases salt tolerance in tomato plants: evidence on the participation of calmodulin-like activities in cross-tolerance signalling. Journal of Experimental Botany **57:** 2391-2400

Clerkx EJM, El-Lithy ME, Vierling E, Ruys GJ, Blankestijin-De Vries H, Groot SPC, Vreugdenhil D, Koornneef M (2004) Analysis of natural allelic variation of Arabidopsis seed germination and seed longevity traits between the accessions Landsberg erecta and Shakdara, using a new recombinant inbred line population. Plant Physiology **135:** 432-443

Collins NC, Tardieu F, Tuberosa R (2008) Quantitative trait loci and crop performance under abiotic stress: where do we stand? Plant Physiology **147:** 469-486

Devicente MC, Tanksley SD (1993) Qtl Analysis of Transgressive Segregation in an Interspecific Tomato Cross. Genetics **134:** 585-596

Dickson MH (1980) Genetic-Aspects of Seed Quality. Hortscience **15:** 771-774

Doerge RW, Churchill GA (1996) Permutation tests for multiple loci affecting a quantitative character. Genetics **142:** 285-294

Dudley JW (1993) Molecular markers in plant improvement - manipulation of genes affecting quantitative traits. Crop Science **33:** 660-668

Ebrahimi A, Maury P, Berger M, Kiani SP, Nabipour A, Shariati F, Grieu P, Sarrafi A (2008) QTL mapping of seed-quality traits in sunflower recombinant inbred lines under different water regimes. Genome **51:** 599-615

El-Kassaby YA (1991) Genetic variation within and among conifer populations: review and evaluation of methods. *In* Fineschi S, Malvolti M.E., Cannata F, H H.H., eds, Biochemical markers in the population genetics of forest trees SPB Academic Publishing bv, The Hague, pp 61-76

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

Finch-Savage WE, Leubner-Metzger G (2006) Seed dormancy and the control of germination. New Phytologist **171:** 501-523

Foolad MR (1996) Response to selection for salt tolerance during germination in tomato seed derived from PI 174263. Journal of the American Society for Horticultural Science **121:** 1006-1011

Foolad MR, Chen FQ (1999) RFLP mapping of QTLs conferring salt tolerance during the vegetative stage in tomato. Theoretical and Applied Genetics **99:** 235-243

Foolad MR, Lin GY (1998) Genetic analysis of low-temperature tolerance during germination in tomato, Lycopersicon esculentum Mill. Plant Breeding **117:** 171-176

Foolad MR, Lin GY (2000) Relationship between cold tolerance during seed germination and vegetative growth in tomato: Germplasm evaluation. Journal of the American Society for Horticultural Science **125:** 679-683

Foolad MR, Lin GY, Chen FQ (1999) Comparison of QTLs for seed germination under non-stress, cold stress and salt stress in tomato. Plant Breeding **118:** 167-173

Foolad MR, Subbiah P, Kramer C, Hargrave G, Lin GY (2003) Genetic relationships among cold, salt and drought tolerance during seed germination in an interspecific cross of tomato. Euphytica **130:** 199-206

Foolad MR, Subbiah P, Zhang L (2007) Common QTL affect the rate of tomato seed germination under different stress and nonstress conditions. International journal of plant genomics **2007:** 97386

Galpaz N, Reymond M (2010) Natural Variation in Arabidopsis thaliana Revealed a Genetic Network Controlling Germination Under Salt Stress. PLoS ONE **5:** e15198

Goodchild NA, Walker MG (1971) A method of measuring seed germination in physiological studies. Annals of Botany **35:** 615-621

Hayashi E, Aoyama N, Still DW (2008) Quantitative trait loci associated with lettuce seed germination under different temperature and light environments. Genome **51:** 928-947

Hilhorst HWM (2007) Definitions and hypotheses of seed dormancy. *In* KJ Bradford, H Nonogaki, eds, Annual Plant Reviews: Seed Development, Dormancy and Germination, Vol 27. Blackwell Publishing, Sheffield, pp 50-71

Hilhorst HWM, Finch-Savage WE, Buitink J, Bolingue W, Leubner-Metzger G (2010) Dormancy in Plant Seeds. *In* E. Lubzens, J Cerda, M Clark, eds, Dormancy and Resistance in Harsh Environments. Springer, Berlin, pp 43-67

Hilhorst HWM, Karssen CM (1988) Dual effect of light on the gibberellin-and nitrate-stimulated seed germination of Sisymbrium officinale and Arabidopsis thaliana. Plant Physiology **86:** 591–597

Jansen RC, Ooijen JW, Stam P, Lister C, Dean C (1995) Genotype-by-environment interaction in genetic mapping of multiple quantitative trait loci. Theoretical and Applied Genetics **91:** 33-37

Johnson WC, Jackson LE, Ochoa O, van Wijk R, Peleman J, St Clair DA, Michelmore RW (2000) Lettuce, a shallow-rooted crop, and Lactuca serriola, its wild progenitor, differ at QTL determining root

architecture and deep soil water exploitation. Theoretical and Applied Genetics **101:** 1066-1073

Joosen RVL, Kodde J, Willems LAJ, Ligterink W, van der Plas LHW, Hilhorst HWM (2010) GERMINATOR: a software package for high-throughput scoring and curve fitting of Arabidopsis seed germination. Plant Journal **62:** 148-159

Joosen RVL, Ligterink W, Hilhorst HWM, Keurentjes JJB (2009) Advances in Genetical Genomics of Plants. Current Genomics **10:** 540-549

Kahn TL, Fender SE, Bray EA, O'Connell MA (1993) Characterization of expression of drought and abscisic acid-regulated tomato genes in the drought-resistant species lycopersicon-pennellii. Plant Physiology **103:** 597-605

Keurentjes JJB, Bentsink L, Alonso-Blanco C, Hanhart CJ, Vries HBD, Effgen S, Vreugdenhil D, Koornneef M (2007) Development of a near-isogenic line population of Arabidopsis thaliana and comparison of mapping power with a recombinant inbred line population. Genetics **175:** 891-905

Kobayashi Y, Koyama H (2002) QTL analysis of Al tolerance in recombinant inbred lines of Arabidopsis thaliana. Plant and Cell Physiology **43:** 1526-1533

Koornneef M, Bentsink L, Hilhorst HWM (2002) Seed dormancy and germination. Current Opinion in Plant Biology **5:** 33-36

Landjeva S, Lohwasser U, Borner A (2010) Genetic mapping within the wheat D genome reveals QTL for germination, seed vigour and longevity, and early seedling growth. Euphytica **171:** 129-143

Langridge P, Paltridge N, Fincher G (2006) Functional genomics of abiotic stress tolerance in cereals. Briefings in Functional Genomics and Proteomics **4:** 343-354

Legendre P, Legendre L (1998) Numerical ecology, Vol 20. Elsevier Science, Amsterdam

Liptay A, Schopfer P (1983) Effect of water-stress, seed coat restraint, and abscisic-acid upon different germination capabilities of 2 tomato lines at low-temperature. Plant Physiology **73:** 935-938

Mackay TFC (2001) The genetic architecture of quantitative traits. Annual Review of Genetics **35:** 303-339

Mackay TFC, Stone EA, Ayroles JF (2009) The genetics of quantitative traits: challenges and prospects. Nature Reviews Genetics **10:** 565-577

Martin B, Tauer CG, Lin RK (1999) Carbon isotope discrimination as a tool to improve water-use efficiency in tomato. Crop Science **39:** 1775-1783

Michel BE, Radcliffe D (1995) A computer-program relating solute potential to solution composition for 5 solutes. Agronomy Journal **87:** 126-130

Ni BR, Bradford KJ (1992) Quantitative models characterizing seed-germination responses to abscisic-acid and osmoticum. Plant Physiology **98:** 1057-1068

Ouyang XR, van Voorthuysen T, Toorop PE, Hilhorst HWM (2002) Seed vigor, aging, and osmopriming affect anion and sugar leakage during imbibition of maize (Zea mays L.) caryopses. International Journal of Plant Sciences **163:** 107-112

Pelgas B, Bousquet J, Meirmans PG, Ritland K, Isabel N (2011) QTL mapping in white spruce: gene maps and genomic regions underlying adaptive traits across pedigrees, years and environments. Bmc Genomics **12:** 145

Perez-Vega E, Paneda A, Rodriguez-Suarez C, Campa A, Giraldez R, Ferreira JJ (2010) Mapping of QTLs for morpho-agronomic and seed quality traits in a RIL population of common bean (Phaseolus vulgaris L.). Theoretical and Applied Genetics **120:** 1367-1380

Quesada V, Garcia-Martinez S, Piqueras P, Ponce MR, Micol JL (2002) Genetic architecture of NaCl tolerance in Arabidopsis. Plant Physiology **130:** 951-963

Ranal MA, Santana DG (2006) How and why to measure the germination process? Revista Brasileira de Botânica **29:** 1-11

Rick CM (1958) The role of natural hybridization in the derivation of cultivated tomatoes of western South America. Economic Botany **12:** 346-367

Rosielle AA, Hamblin J (1981) Theoretical aspects of selection for yield in stress and non-stress environments. Crop Science **21:** 943-946

Spano C, Buselli R, Castiglione MR, Bottega S, Grilli I (2007) RNases and nucleases in embryos and endosperms from naturally aged

wheat seeds stored in different conditions. Journal of Plant Physiology **164:** 487-495

Tanksley SD (1993) Mapping polygenes. Annual Review of Genetics **27:** 205-233

Thomson AJ, El-Kassaby YA (1993) Interpretation of seed-germination parameters. New forests **7:** 123-132

Vallejo AJ, Yanovsky MJ, Botto JF (2010) Germination variation in Arabidopsis thaliana accessions under moderate osmotic and salt stresses. Annals of Botany **106:** 833-842

Van Ooijen JW (1999) LOD significance thresholds for QTL analysis in experimental populations of diploid species. Heredity **83:** 613-624

Van Ooijen JW, Maliepaard C (2003) MapQTL®, Version 5.0: Software for the Calculation of QTL Positions on Genetic Maps. *In*. Institute of Plant Genetics, Polish Academy of Sciences, p 305

Van Ooijen JW, Voorrips RE (2001) JoinMap® 3.0, Software for the calculation of genetic linkage maps. Plant Research International, Wageningen, The Netherlands**:** 1–51

Voorrips RE, Verkerke W, Finkers R, Jongerius R, Kanne J (2000) Inheritance of taste components in tomato. Acta Physiologiae Plantarum **22:** 259-261

Wahid A, Gelani S, Ashraf M, Foolad MR (2007) Heat tolerance in plants: an overview. Environmental and Experimental Botany **61:** 199-223

Wudiri BB, Henderson DW (1985) Effects of water-stress on flowering and fruit-set in processing-tomatoes. Scientia Horticulturae **27:** 189-198

Zhang J, Zheng HG, Aarti A, Pantuwan G, Nguyen TT, Tripathy JN, Sarial AK, Robin S, Babu RC, Nguyen BD, Sarkarung S, Blum A, Nguyen HT (2001) Locating genomic regions associated with components of drought resistance in rice: comparative mapping within and across species. Theoretical and Applied Genetics **103:** 19-29

Zhu JK (2002) Salt and drought stress signal transduction in plants. Plant Biology **53:** 247--273

Chapter 3

NATURAL VARIATION FOR SEEDLING TRAITS AND THEIR LINK WITH SEED DIMENSIONS IN TOMATO

Published in PLoS One
Khan N*, Kazmi RH*, Willems LAJ, van Heusden AW, Ligterink W, Hilhorst HWM (2012). PLoS One 7: e43991

*Equal contribution

Abstract

The success of germination, growth and final yield of every crop depends to a large extent on the quality of the seeds used to grow the crop. Seed quality is defined as the viability and vigor attribute of a seed that enables the emergence and establishment of normal seedlings under a wide range of environments. We attempt to dissect the mechanisms involved in the acquisition of seed quality, through a combined approach of physiology and genetics. To achieve this goal we explored the genetic variation found in a RIL population of *Solanum lycopersicum* (cv. Moneymaker) x *Solanum pimpinellifolium* through extensive phenotyping of seed and seedling traits under both normal and nutrient stress conditions and root system architecture (RSA) traits under optimal conditions. We have identified 62 major QTLs on 21 different positions for seed, seedling and RSA traits in this population. We identified QTLs that were common across both conditions, as well as specific to stress conditions. Most of the QTLs identified for seedling traits co-located with seed size and seed

weight QTLs and the positive alleles were mostly contributed by the *S. lycopersicum* parent. Co-location of QTLs for different traits might suggest that the same locus has pleiotropic effects on multiple traits due to a common mechanistic basis. We show that seed weight has a strong effect on seedling vigor and these results are of great importance for the isolation of the corresponding genes and elucidation of the underlying mechanisms.

Introduction

The success of germination, seedling establishment and later growth and development of every agricultural crop depends on many factors. Among the various factors seed quality is one of the most important factor to affect the success of crops (Finch-Savage, 1995). High quality seed is a composite term used for all the attributes that add to the performance of a seed: genetically and physically pure, vigorous, viable, a high rate of germination, free from seed borne diseases and heat damage and produce normal seedlings under various environmental (stress) conditions (Dickson, 1980; Hilhorst and Toorop, 1997; Hilhorst and Koornneef, 2007). Seed quality is also drastically affected by various environmental conditions during seed development, as well as subsequent harvesting methods, handling, and storage conditions. All these environmental factors interact with the seed's genetic make-up (Coolbear, 1995; McDonald, 1998; Koornneef et al., 2002).

Good seedling establishment and seedling vigor are essential for sustainable and profitable crop production and is therefore considered the most critical stage of a developing crop. Low seed vigor greatly influences both the number of emerging seedlings, and the timing and uniformity of seedling emergence. This has a major impact upon many aspects of crop production that determine cost effectiveness and the inputs required, and

also has direct influence on the yield and marketing quality of a crop (Bleasdale, 1967; Finch-Savage, 1995) and subsequent efforts or amount of inputs during later stages of crop development will not compensate for this upshot. In tomato, huge phenotypic variation has been observed among the seeds of different species. The seeds of cultivated tomato have developed to be several times larger than their wild counterparts as a result of domestication and breeding (Doganlar et al., 2000). A number of QTL studies carried out on several populations of interspecific crosses between cultivated tomato and their wild relatives have allowed the identification of loci controlling seed weight (Tanksley et al., 1982; Weller et al., 1988; Goldman et al., 1995; Grandillo and Tanksley, 1996). Seed weight is an indication of the reserves that seeds contain and large and heavy seeds reveal that the seed has more reserved food (Wright and Westoby, 1999). Many studies have shown that initial seedling size is positively related to seed size, and larger seeds have better seedling survival rate as well as higher competitiveness both within species (Dolan, 1984; Morse and Schmitt, 1985; Wulff, 1986; Winn, 1988; Tripathi and Khan, 1990; Wood and Morris, 1990; Zhang and Maun, 1991; Moegenburg, 1996) and among species (Stebbins, 1976; Stanton, 1984; Morse and Schmitt, 1985; Marshall, 1986; Winn, 1988; Tripathi and Khan, 1990; Wood and Morris, 1990; Seiwa and Kikuzawa, 1991; Jurado and Westoby, 1992; Chambers, 1995; Seiwa and Kikuzawa, 1996; Greene and Johnson, 1998; Cornelissen, 1999). The seed supplies the embryo with sufficient nutrition and energy during germination from the food reserves that the seed acquires during the seed filling phase. Thus the seed filling phase plays a crucial role in successful establishment of an autotrophically growing seedling by supplying nutrition and energy and bridging the gap between germination and establishment of green cotyledons that are capable of photosynthesis

(Ellis, 1992; Castro et al., 2006).

Root systems perform the crucial task of providing water, nutrients and physical support to the plant. The length of the main root and the density of the lateral roots determine the architecture of the root system in tomato and other dicots and play a major role in determining whether a plant will succeed in a particular environment (Malamy and Benfey, 1997). Seed size may have an essential role in improvement of root architecture during its initial downward growth (Jurado and Westoby, 1992). Dissecting natural variation in seed vigor of *Brassica oleracea* Finch-Savage *et al.*, (2010) found a strong effect of seed vigor on the initial downward growth of seedlings and fine mapped QTLs for rapid initial growth of root which co-located with seed weight QTLs.

Little is known about the role of tomato seed size in seedling growth. In tomato, seed germination and early seedling growth are the most sensitive stages to environmental stresses such as salinity, drought and extreme temperatures (Jones, 1986) and most of the cultivated tomatoes are considered to be sensitive to abiotic stress conditions (Maas, 1986; Foolad et al., 1997; Foolad et al., 1998). Considerable genetic variation for abiotic stress tolerance exists within cultivated tomato (*Solanum lycopersicum)*, as well as in its related wild species such as *S. habrochaitis, S. pimpinellifolium*, and *S. pennellii* (Cannon et al., 1973; Scott and Jones, 1982; Wudiri and Henderson, 1985; Wolf et al., 1986). The wild type tomato germplasm is a rich source of desirable genetic variability and many wild species have been identified with higher tolerance to abiotic stresses (Rick, 1973, 1982; Foolad et al., 2007). Among the wild species of tomato, S*. pimpinellifolium* provides numerous benefits for studying the natural genetic variation and morphological characters. It is amenable to experimental culture, readily hybridized, quick-growing, highly

reproductive, relatively well known genetically and relatively resistant to biotic and abiotic stress (Stubbe, 1960, 1965; Rick et al., 1977; Foolad et al., 2007) and it is closely related to *S. lycopersicum*. Despite their close relationship, the two species differ greatly in many morphological and economically interesting traits, not only in fruit size and growth traits (Rick, 1958; Grandillo and Tanksley, 1996), but also in seed size (Grandillo and Tanksley, 1996; Doganlar et al., 2000; Doganlar et al., 2002).

In general, seed and seedling vigor characteristics are complex traits, which are probably controlled by several genes and are therefore suitable for quantitative trait loci (QTL) analysis. In the current study we analyzed these traits in a recombinant inbred line (RIL) population between *S. lycopersicum* (cv. Money maker) and *S. pimpinellifolium*(Voorrips et al., 2000; Kazmi et al., 2012). The study revealed the presence of high phenotypic variability in the population with regard to seed size, seedling growth and root architecture and due to this variability we were able to identify 62 QTLs related to seed and seedling traits. In addition the results also revealed a strong correlation between seed size and seedling growth and co-location of QTLs for these traits.

Materials and Methods

Plant material

The tomato RIL population was obtained from a cross between *Solanumlycopersicum* cv. Moneymaker and *Solanumpimpinellifolium* CGN 15528 (Voorrips et al., 2000). This population was genotyped for a total of 865 Single Nucleotide Polymorphism (SNP) markers in F_7 and produced 83 RILs in the F_8. The genotyping was done with a custom made, in house SNP array based on polymorphisms detected with 454 (Roche) and Illumina sequencing in 8 different tomato species (personal communication AW van Heusden).

Growth conditions and seed collection

The RIL population of *S. lycopersicum* X *S. pimpinellifolium* was grown twice under controlled conditions in the greenhouse facilities at Wageningen University, the Netherlands. The day and night temperatures were maintained at 25 and 15 °C, respectively, with 16 h light and 8 h dark (long-day conditions). All the RILs were uniformly supplied with the basic dose of fertilizer.

Seeds were collected from healthy mature fruits and subsequently treated with 1% hydrochloric acid (HCL) for 1.5 h to remove the pulp sticking onto the seeds. The solution of tomato seed extract with diluted hydrochloric acid was passed through a fine mesh sieve and washed with tap water to remove pulp and hydrochloric acid. The seeds were processed and disinfected by soaking in a solution of trisodium phosphate (Na3PO4.12H2O). Finally, seeds were dried on filter paper at room temperature and were brushed to remove impurities with a seed brusher (Seed Processing Holland BV, Enkhuizen, The Netherlands, http://www.seedprocessing.nl). The cleaned seeds were dried for 3 d at 20°C and stored in a storage room (13 °C and 30% RH) in paper bags. The seeds of each harvest were bulked separately for each RIL and were used in the subsequent experiments.

Linkage analysis

The genetic linkage map consists of 12 individual linkage groups corresponding to the 12 chromosomes of tomato and was made on the basis of genotyping the segregation of parental alleles in the *S. lycopersicum* cv. Moneymaker X *S. pimpinellifolium* G1.1554 RIL population with 865 SNP markers. See Kazmi *et al.*, 2012 for more details.

Phenotyping of seed traits of the RIL population

Seed weight (SW) was measured as the average seed weight of a batch of 100 seeds. Seed size was determined by taking close-up photographs from 2 x 100 seeds using a Nikon D80 camera with a 60mm objective fixed to a repro stand and connected to a computer, using Nikon camera control pro software version 2.0 (Joosen et al., 2010). The photographs were analyzed using the open source image analysis suite ImageJ (http://rsbweb.nih.gov/ij/) by using color-thresholds combined with particle analysis that automatically scored seed size (SS) as the area of selection in square pixels, circularity (SC) as $4\pi*(\text{area}/\text{perimeter}^2)$and seed length (SL) as the longest distance between any two points along the selection boundary (feret's diameter). Seed size and seed length was also determined in 12-h imbibed seeds (ImbSS and ImbSL, respectively).

Seedling growth

Seedling growth was tested in three independent experiments. In the first two experiments seedlings were grown on vertical plates (12 x 12 cm square Petri dishes) on half MS medium under aseptic conditions at pH 5.6. The top 4 cm of the agar solution was removed with a sterilized knife and the seedlings were grown on the remaining 8 cm. In each experiment 7 seedlings were grown per plate in a randomized complete block design for each harvest in duplicate (7*2*2 seedlings per experiment) in a climate chamber at 25 °C with long day conditions (16h light, 8h dark). Before sowing, seeds were surface sterilized for 16h in a desiccator over a solution of 100 ml 4% sodium hypochlorite + 3 ml concentrated hydrochloric acid.

Germination was scored at 8-h intervals as visible radical protrusion. After the start of germination photographs were taken at 24–h intervals for root architecture analysis. Five days after germination the hypocotyl length

and the fresh root and shoot weight data were measured (HypL, FrRt and FrSh respectively). After subsequent drying for 1 week at 90 ^{0}C the dry root and shoot weights were measured (DrRt and DrSh respectively). Root system architecture was analyzed with the EZ-Rhizo software package (Armengaud et al., 2009) to obtain parameters such as total root size (TRS), main root length after five days (MRL), number of lateral roots per main root (LRn) and lateral root density per branch zone (LRD-Bz).

In a third experiment seedlings were grown under nutrient-deprived conditions on a Copenhagen table. The seedlings were grown on blue filter paper and were covered with conical glasses with a small hole on the top. These conical glasses prevent the loss of moisture provided by the Copenhagen table without blocking aeration of the seedlings. Each harvest was tested separately in two consecutive sub-sets of experiments. Twenty seeds of each RIL for each seed harvest were germinated on Copenhagen tables in a randomized complete block design in triplicate (20x3x2 harvests). Germination was recorded as visible radical protrusion at 8-h intervals. The first 10 germinated seeds were allowed to develop into a seedling and ten days after reaching the t_{50} (time to 50 percent germination) the seedlings were harvested and the fresh and dry root and shoot weight data were determined (FrRtwn, DrRtwn, FrShwn and DrShwn, respectively). In this case we could not assess the root architecture due to the set-up of the Copenhagen table on which the roots grow horizontally and become intertwined.

Data analysis

Pearson correlations between different traits were calculated with the PASW statistics software, version 17 (Arbuckle, 1999). QTL analyses was performed with the mapping software MapQTL®5.0 (Van Ooijen and

Maliepaard, 2003). In a first step, putative QTLs were identified using interval mapping. Thereafter, the estimated additive effect and the percentage variance explained by each QTL, as well as the total variance explained by all of the QTLs affecting a trait, were obtained by MQM mapping. For this purpose different markers were tested around a putative QTL position as a cofactor (Van Ooijen and Maliepaard, 1996) and those maximizing the LOD score were selected as the final cofactors and finally restricted multiple QTL mapping (rMQM) was used to obtain the confidence intervals. A LOD score of 2 was calculated as a threshold level with a permutation test to detect statistically significant QTL.

Analysis of heritability and epistasis

Broad-sense heritability (h^2_b) was estimated from one-way random-effects of analysis of the variance (ANOVA, SPSS version 19.0) with the equation: $h^2_b = \sigma^2_g / (\sigma^2_g + \sigma^2_e)$ where σ^2_g is the genetic variance and σ^2_e is the environmental variance (Keurentjes et al., 2007). Significant differences among all means of the RILs were estimated using one-way ANOVA followed by a least significant difference (LSD) test.

A two-dimensional genome-wide epistatic interactions analysis was performed using the R/qtl software package (Broman et al., 2003) in order to identify epistatic interactions contributing to variation in traits. This includes nested linear model-fitting for each pair of loci (Koller et al., 2009). Genome-wide significance thresholds were obtained by 10,000 permutation tests (Doerge and Churchill, 1996) with the Haley-Knott regression method (Broman et al., 2003). LOD significance threshold of the maximum genome-wide interaction (lod.int), full model (lod.full), and conditional interactive model (lod.fv) were found to be 4.09, 6.04 and 4.63, respectively.

Results

Phenotypic variation in seed and seedling vigor related traits.

In total 19 traits were tested in this study, including 6 seed traits, such as seed weight (SW), seed size (SS), seed length (SL), seed circularity (SC), imbibed seed size (ImbSS) imbibed seed length (ImbSL) and 5 seedling- and 4 root architecture related traits. The seedling related traits included fresh and dry root and shoot weight (FrRt, DrRt, FrSh and DrSh respectively), and hypocotyl length (HypL). The 4 root architecture related traits, included main root path length (MRL), total root size (TRS), lateral root number (LRn), and lateral root density per branched zone (LRD/Bz) in both experiments. Differences between the two parents were statistically highly significant for all the traits studied ($P < 0.01$ to 0.001) with the *S. lycopersicum* parent having higher trait values as compared to the *S. pimpinellifolium* parent in all the traits except LRD/Bz (Table 3.1). In addition, there were statistically significant differences for these traits among the different lines of the RIL population (Table 3.1).

Besides testing on agar plates, we measured seedling growth of the RIL population also on a Copenhagen table without any nutrition, to test the importance of amount of reserve food present in the seed (seed vigor) in the form of total biomass acquired by the seedling in a specific period of time from radical protrusion until harvesting of the seedling. In this experiment we measured fresh and dry root and shoot weight (FrRtwn, DrRtwn, FrShwn and DrShwn respectively). We observed significant differences between the two parents as well as in the RIL population for the seedling traits measured during this experiment (Table 3.1). There was 27 to 56 % decrease in the biomass gained in ten days after germination under the nutrientless condition as compared to the mass obtained in five

Table 3.1. Phenotypic analysis of seed and seedling related vigor traits of a *S. lycopersicum* and *S. pimpinellifolium* RIL population and its two parents

Nr	Trait[1]	S. lycopersicum	S. pimpinellifolium	RIL Population			F-Value[3]	P-Value[3]
		Mean	Mean	Mean		SD[2]		
1	FrRt	20.30	10.9	15.91	±	5.21	3.58	0.001
2	DrRt	1.97	0.56	1.19	±	0.36	2.13	0.001
3	FrSh	46.27	17.01	32.47	±	8.97	4.51	0.001
4	DrSh	3.04	1.18	2.18	±	0.50	4.50	0.001
5	HypL	3.20	2.08	2.83	±	0.61	4.00	0.001
6	SW	2.95	1.08	1.70	±	0.38	2.76	0.001
7	SS	4.4	2.34	3.26	±	0.50	16.35	0.001
8	SL	2.93	1.62	2.51	±	0.21	1.56	0.012
9	ImbSS	6.45	3.42	4.72	±	0.75	14.52	0.001
10	ImSL	3.79	2.01	3.08	±	0.25	1.39	0.046
11	FrShwn	27.20	7.28	13.37	±	3.54	8.27	0.001
12	DrShwn	1.47	0.37	0.77	±	0.20	7.20	0.0001
13	FrRtwn	14.64	5.48	9.06	±	2.52	10.89	0.001
14	DrRtwn	0.95	0.31	0.52	±	0.15	2.96	0.001
15	MRL	8.54	4.61	6.93	±	1.18	3.47	0.001
16	TRS	13.99	6.36	10.18	±	2.38	3.53	0.001
17	LRn	8.60	3.86	4.65	±	2.15	3.57	0.001
18	LRD/BZ	3.41	6.08	4.65	±	2.90	1.15	0.245

[1]FrRt = Fresh Root weight, FrSh = Fresh shoot weight, DrRt = Dry root weight. DrSh = Dry Shoot weight, HypL = Hypocotyl length, SW= Dry Seed weight. SS = Dry seed size, SL = Dry seed length, SC =Dry Seed circularity, ImbSS = imbibed seed size, ImbSL = Imbibed seed length, FrShwn = Fresh Shoot weight under nutrientless condition, DrShwn =Dry shoot weight in nutrientless condition, FrRtwn = Fresh root weight in nutrientless condition, DrRtwn = Dry root weight under nutreintless condition, MRL =Main Root path Length, TRS = Total root size, LRn =

Lateral root number per main root, LRD/Bz = Lateral roots density per branched zone.

[2]standard deviation.

[3]F-value and P- value were calculated for the population mean.

days after germination under the normal nutrient conditions (Table 3.2). All measured traits showed a normal distribution over the RIL population (Figure 3.1). Figure 3.1 also shows that transgression was present for most traits.

Table3.2. Reduction in biomass of seedling grown under nutrient stress condition as compared to the biomass obtained under normal nutrient conditions.

Normal[1]				Wn[2]				Decr[3]		
Trait[4]	*S. lyco*	*S. pimp*	RILs	Traits[5]	*S .lyco*	*S. pimp*	RILs	*S. lyco*	*S. pimp*	RILs
	Mean	Mean	Mean		Mean	Mean	Mean			
FrRt	20.3	10.9	15.9	FrRtwn	14.64	5.48	9.06	27.90%	49.70%	43.10%
DrRt	1.97	0.56	1.19	DrRtwn	0.95	0.31	0.52	51.80%	44.60%	56.30%
FrSh	46.27	17.01	32.5	FrShwn	27.21	7.28	13.3	41.20%	57.20%	58.80%
DrSh	3.04	1.18	2.18	DrShwn	1.47	0.37	0.77	51.60%	68.70%	64.70%

[1]Normal = Seedling grown under normal nutrients condition,

[2]Wn= Seedling grown on Copenhagen table without nutrition,

[3]Decr = Seedling grown on Copenhagen table without nutrition,

[4]FrRt = Fresh Root weight, DrRt = Dry root weight, FrSh = Fresh shoot weight, DrSh = Dry Shoot weight,

[5]FrRtwn = Fresh root weight in nutrientless condition, DrRtwn = Dry root weight under nutrientless condition, FrShwn = Fresh Shoot weight under nutrientless condition, DrShwn = Dry shoot weight in nutrientless condition.

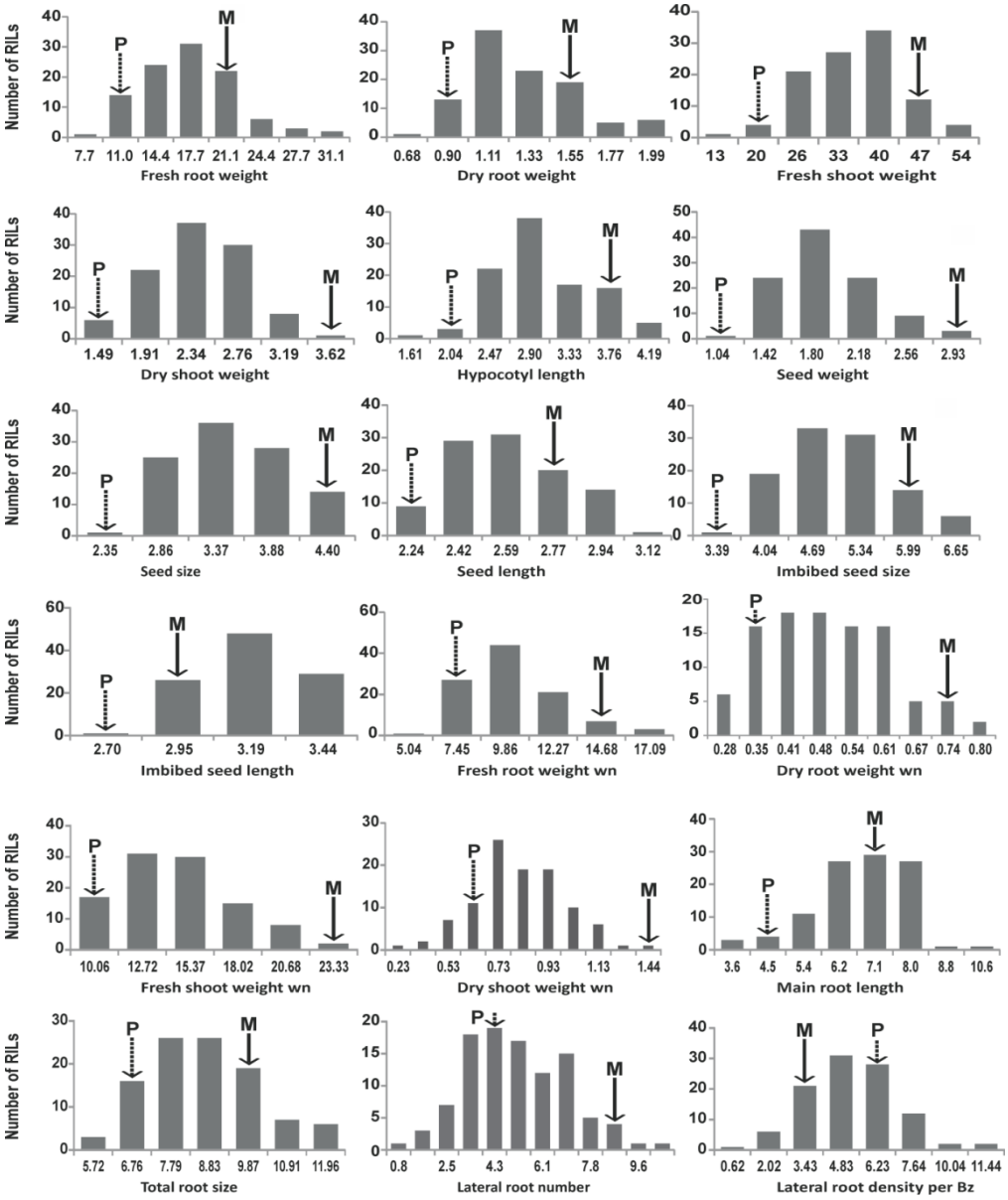


Figure 1. Frequency distributions of non-normalized data of all measured seed and seedling phenotypes in the *Solanum lycopersicum* x *Solanum pimpinellifolium* RIL population. wn: without nutrition. The parental values are indicated with a solid arrow. P = *S. pimpinellifolium* parent and M= *S. lycopersicum* parent.

Correlation between traits

Statistically significant correlations were observed between seed weight and seedling traits such as fresh and dry root and shoot weight (Figure3.2). The R^2 value for the Pearson correlation between seed weight and different seedling traits varied from 0.64 for seed weight vs. fresh root weight to 0.78 for seed weight vs. dry shoot weight (Figure3.2). Under the nutrient-deprived condition the R^2 value varied from 0.58 to 0.83 between seed weight and dry root and shoot weight (DrRtwn and DrShwn). In addition, we found statistically significant correlations among seed traits such as seed size and seed length and seedling traits, as expected (data not shown). On the other hand, although we found significantly negative correlation between seed size and seed circularity, we found no correlations between seed circularity and seedling traits. In case of root architecture, we found low (R^2 value 0.44 and 0.45), but statistically highly significant (p value 0.001) correlations between seed weight and total root size (TRS) and lateral root number(LRn), but could not find any correlation with the other root traits (MRL and LRD/Bz)(Figure 3.2). We also tested the correlation between seed traits and seed performance such as total germination percentage (G_{max}%), rate of germination (t_{50}) and uniformity of germination (U_{7525}) (Kazmi et al., 2012), but found no significant correlations between seed traits and seed germination parameters, which is obvious from the R^2 values (Figure 3.2).

Mapping QTLs for different traits

We used the data of the studied seed, seedling and RSA phenotypes under control and nutrient-deprived conditions to map QTLs

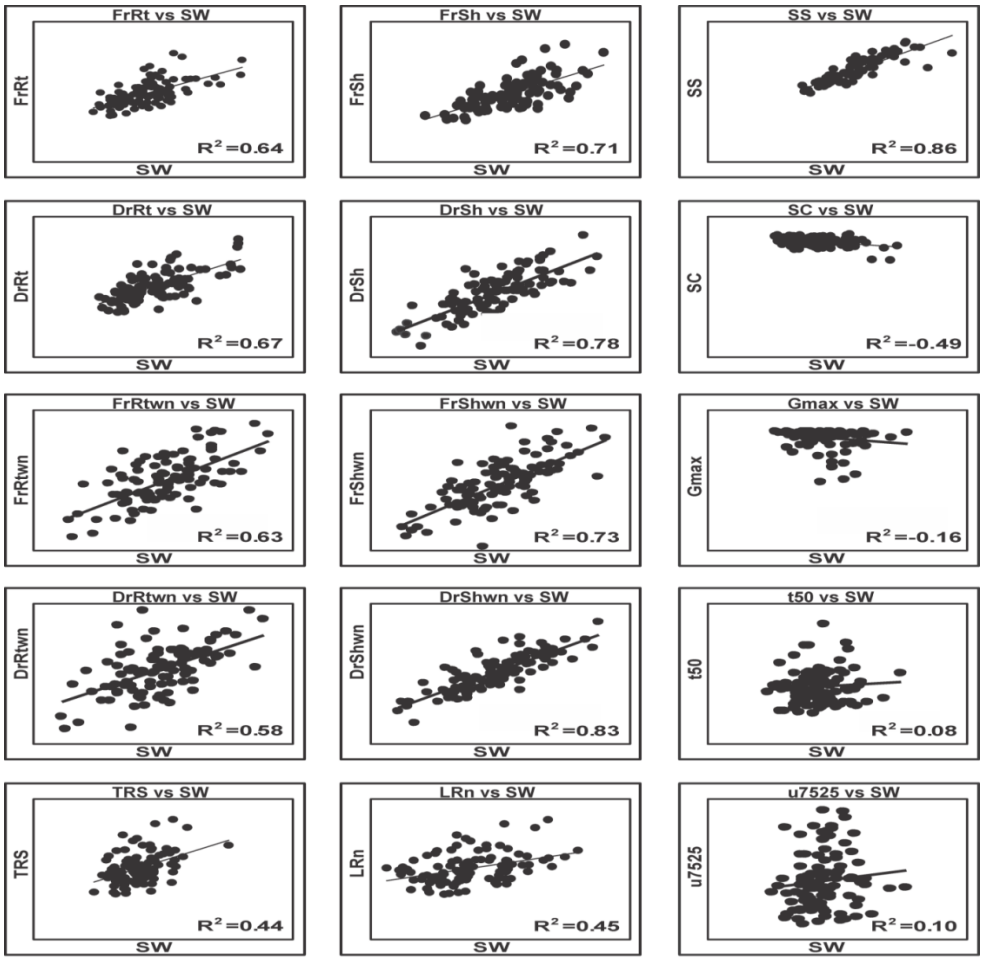


Figure 3.2. Correlation among seed and seedling traits. SW = Seed weight, SS = Seed size, SL = Seed length, FrRt = Fresh root weight, DrRt = Dry root weight, FrSh = Fresh shoot weight, DrSh = Dry shoot weight, FrShwn = Fresh shoot weight in nutrientless conditions, DrShwn = Dry shoot weight in nutrientless conditions, FrRtwn = Fresh root weight in nutrientless conditions, DrRtwn = Dry Root weight in nutrientless conditions, G_{max} = Maximum total germination in %, t_{50} = time to complete 50% germination, U_{7525} = Uniformity of germination (time between 25 to 75% germination).

Table 3.3. Overview of significant QTLs associated with seed and seedling traits of *S. lycopersicum* and *S. pimpinellifolium* tomato RIL population.

Trait[1]	Chr[2]	Confidence Interval (cM)	Nearest Marker[3]	LODscore	Additive Effect[4]	Explained Variance (%)	Total Explained Variance (%)	Heritability
FrRt								
	9	54.8-91.8	62162316	3.3	-0.73	14.1	30.8	0.78
	10	8.6-100.4	58738936	2.1	0.59	8.5		
	12	0.0-79.8	62040100	2.0	-0.55	8.2		
DrRt								
	9	46.7-101.1	60488088	2.6	-0.70	11.9	11.9	0.68
FrSh								
	9	59.0-96.3	62897108	3.4	-0.78	16	16.0	0.82
DrSh								
	4	0.0-20.9	30398	2.6	0.63	9.7	25.1	0.82
	9	65.0-88.5	62897108	3.4	-0.75	15.4		
HypL								
	1	18.9-64.9	2766897	2.0	-0.54	7.4	33.7	0.80
	6	87.3-99.2	41812268	4.2	-0.84	17		
	10	1.6-80.2	59476312	2.4	-0.69	9.3		
SW								
	1	49.9-64.9	69227784	3.1	-0.56	8.6	60.9	0.73
	4	50.4-63.8	51677496	4.6	-0.69	13.5		
	6	95.8-109.3	44905196	3.1	0.57	8.5		
	9	54.8-95.3	60488088	4.2	-0.68	12.1		
	9	54.2-94.3	64960580	3.6	-0.63	8.4		
	11	0.0-28.5	4775141	3.7	-0.62	9.8		
SS								
	1	44.8-64.9	69430752	2.2	-0.49	7.0	36.5	0.94
	4	49.4-67.7	51677496	3.7	-0.64	12.1		
	9	52.3-104.1	64960580	2.6	-0.53	8.2		
	11	0.0-20.6	5148394	2.9	-0.56	9.2		
SL								
	2	0.0-92.3	39990428	3.2	0.83	9.1	33.3	0.61
	9	0.0-35.8	48774	2.4	-0.56	8.0		
	11	22.1-33.5	48283252	4.6	-0.73	16.2		

Table 3.3 *Continued*

Trait[1]	Chr[2]	Confidence Interval (cM)	Nearest Marker[3]	LODscore	Additive Effect[4]	Explained Variance (%)	Total Explained Variance (%)	Heritability
SC								
	3	85.7-135.2	58802824	3.0	0.64	8.1	51.9	0.70
	4	0.0-74.1	3902301	2.0	0.50	5.4		
	6	86.3-104.3	42299156	3.9	-0.70	11.1		
	8	79.3-124.4	57594496	2.6	0.56	7		
	9	0.0-16.7	1751657	4.4	0.75	12.6		
	11	20.6-52.1	48283252	2.8	0.57	7.7		
ImbSS								
	4	46.0-69.2	51677496	2.6	-0.59	9.3	41.3	0.93
	6	58.5-109.3	43431568	2.2	0.53	7.6		
	9	56.0-93.0	64960580	3.0	-0.65	10.9		
	11	0.0-16.0	5148394	3.7	-0.72	13.5		
ImbSL								
	9	28.5-63.5	5400867	2.6	-0.68	10.6	21.3	0.58
	11	0.0-36.4	5472482	2.3	-0.65	10.7		
FrRtwn								
	1	20.5-36.3	2746777	3.6	-0.66	11.7	45.2	0.89
	6	36.6-81.6	39180864	3.0	0.59	9.5		
	7	64.3-90.7	61282892	2.0	-0.48	6.5		
	9	81.3-95.3	64960580	3.1	-0.60	10.0		
	11	0.0-68.4	4775141	2.4	-0.52	7.5		
DrRtwn								
	6	43.6-80.5	37874180	2.1	0.64	9.9	23.6	0.88
	9	46.7-95.3	62897108	2.9	-0.71	13.7		
FrShwn								
	1	57.9-64.9	69430752	6.4	-1.01	24.6	36.1	0.92
	9	76.4-96.3	64960580	3.3	-0.69	11.5		
DrShwn								
	9	70.3-96.3	64960580	3.2	-0.78	14.6	14.6	0.75
MRL								
	1	1.0-39.5	2746777	2.6	-0.51	6.1	41.3	0.65
	2	29.4-67.8	37722740	2.5	0.59	6.0		

Table 3.3 *continued*

Trait[1]	Chr[2]	Confidence Interval (cM)	Nearest Marker[3]	LODscore	Additive Effect[4]		Total Explained Variance (%)	Heritability
TRS								
	7	33.2-55.3	28075704	2.7	0.53	6.5		
	9	26.4-104.7	62162316	3.5	-0.63	8.5		
	9	76.4-98.8	65815200	5.7	-0.87	14.2		
	1	0.0-39.5	2746777	2.1	-0.49	5.6	51.4	0.79
	3	59.7-135.2	61881752	2.2	-0.53	5.9		
	9	39.4-75.1	60488088	4.1	-0.70	11.3		
	9	77.4-101.1	65815200	5.6	-0.86	15.7		
	10	9.3-82.2	58738936	2.1	-0.46	4.8		
	11	0.0-12.1	4106782	3.0	-0.60	8.1		
LRn								
	5	53.4-86.1	6814273	2.9	0.71	13.0	32.1	0.78
	11	2.4-22.7	5148394	4.1	-0.87	19.1		
LRD/Bz								
	2	50.0-83.8	43635344	2.6	-0.70	9.4	44.9	0.53
	7	29.2-56.3	3317484	3.8	-0.81	14.5		
	8	22.2-98.9	2908496	2.5	0.64	9.3		
	9	33.8-88.7	62162316	3.2	0.69	11.7		

[1]FrRt = Fresh Root weight, DrRt = Dry root weight, FrSh = Fresh shoot weight, DrSh = Dry Shoot weight, HypL = Hypocotyl length, SW= Dry Seed weight. SS = Dry seed

size, SL = Dry seed length, SC = Dry Seed circularity, ImbSS = Imbibed seed size, ImbSL = Imbibed seed length, FrShwn = Fresh Shoot weight under nutrientless condition,

FrRtwn = Fresh Root weight under nutrientless condition, DrShwn = Dry shoot weight under nutrientless condition, DrRtwn =Dry root weight under nutrientless condition, MRL =Main Root Length, TRS = Total root size, LRn = Lateral root number per main root, LRD/Bz = Lateral roots density per branched zone.

[2]Chromosome on which the QTLs were detected.

[3]Nearest marker to the position of the identified QTLs.

[4]A positive sign means that the allele of *S. pimpinellifolium* contributed to the increase of particular trait while the negative sign means that the allele of *S. lycopersicum* increased the trait at this particular locus.

with the use of a LOD threshold of 2.0. Multiple QTL (MQM) mapping analysis revealed a total of 62 significant QTLs on 21 different positions for the 19 seed and seedling traits tested across the RIL population (Table 3.3). By making a heat map of LOD profiles, QTLs can be visualized and global 'hot spots' and empty regions across the 12 chromosomes can be seen (Figure3.3). Co-localization of QTLs was found for different seed and seedling traits on the bottom of chromosomes 1, 4, 6, 9 and 11 (Table 3.3, Figure 3.3). Out of the 62 detected QTLs, 25 were related to seed traits, such as seed weight, seed size, seed length and seed circularity. Seventeen QTLs were related to seedlings biomass , such as fresh and dry root and shoot weight (across both the growing conditions) and 3 QTLs to hypocotyl length, whereas 17 QTLs were related to root system architecture. We identified significant QTLs for all the traits, ranging from 1 to 6 QTLs per trait with LOD scores in the range of 2.1 to 6.4. Explained variances for single QTL ranged from 4.8% for the QTL for total root size on chromosome 10 to 24.6% for the QTL on chromosome 1 for fresh shoot weight without nutrition. The total explained variance for different traits caused by these QTLs varied from11.9 % for dry root weight to 62.9 % for seed weight with genetic heritability ranging from 0.53 for lateral root density to 0.94 for seed size. About 72.5% of the favorable alleles were derived from the *S. lycopersicum* parent (negative additive effects in Table 3.3).

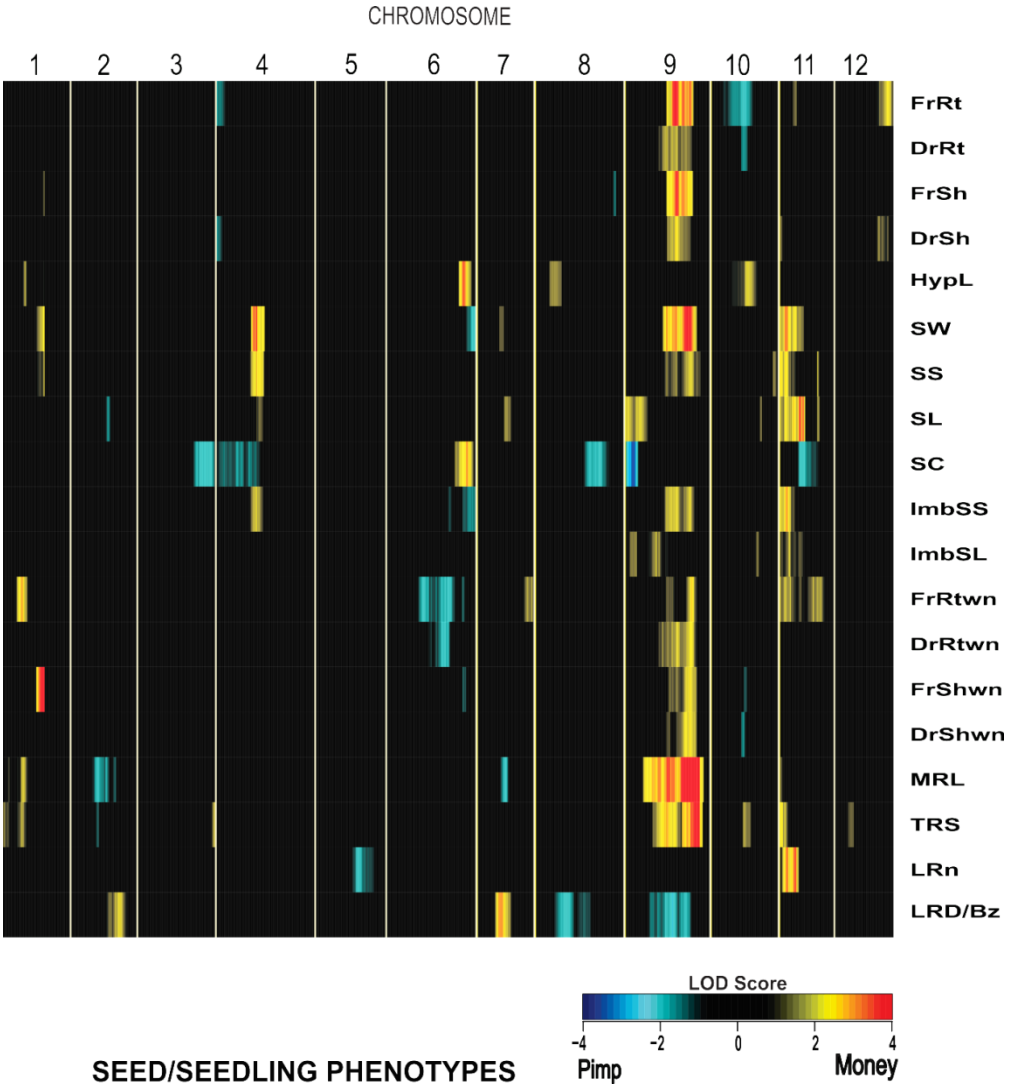


Figure 3.3 Heatmap of QTLs identified for seed and seedling quality traits. Tomato chromosomes are identified by numbers (1–12), with centimorgans ascending from the left to right; chromosomes are separated by yellow lines. SW=Seed Weight, SS=Seed Size SL=Seed Length. FrRt= Fresh Root weight, DrRt= Dry Root weight, FrSh = Fresh Shoot weight, DrSh Dry Shoot weight, FrShwn =Fresh Shoot weight in nutrientless conditions, DrShWn =Dry Shoot weight in nutrientless

conditions, FrRtwn =Fresh Root weight in nutrientless conditions, DrRtwn =Dry Root weight in nutrientless conditions, MRL =Main Root Length, TRS =Total Root Size, LRn =Lateral Root number per main root, LRD/Bz =Lateral Root Density per Branched zone. Colors indicate QTLs significant at $P = 0.002$ in multiple QTL mapping models (1-LOD intervals). Blue and light blue colors indicate a larger effect of the trait in *S. pimpinellifolium,* and yellow and red in *S. lycopersicum.*

Stress specific QTLs

We identified QTLs that were either common across both the conditions or specific to a particular condition. For example the QTLs on chromosome 9 could be identified for the 4 seedling traits tested across both the conditions while the QTLs on chromosomes 4 and 12 for FrRt could only be identified under normal nutrient conditions (Table 3.3, Figure 3.3). On the other hand the QTLs on Chromosome 1 for FrRtwn, and FrShwn and on Chromosome 6 for FrRtwn and DrRtwn, as well as on chromosome 7 and 11 for FrRtwn were only identified under nutrient-deprived conditions.

Epistatic Interactions

For each of the described traits, a genome-wide epistasis analysis was performed. In this analysis all pairwise combinations of the markers closest to each target QTL was tested. With this method several instances of epistatic interactions among seed size and seedling QTLs were revealed (Table 3.4, Figure3.4).

These epistatic interactions contribute to phenotypic variability, but hinder detection and affect estimation of QTLs examined singly. This analysis revealed novel loci on several chromosomes interacting to

influence seed size and seedling traits. The analysis revealed loci on chromosomes 8 and 11 interacting to influence seed circularity (Table 3.4, Figure 3.4). Similarly, for seed length, evidence of interaction was observed on chromosomes 4 and 7. A two-way interaction was also revealed for total root size on chromosomes 9 and 11. Finally, a strong interaction was observed for lateral root density between a locus on chromosome 7 and 8 (LODint =6.97) (Table 3.4, Figure 3.4), which had the highest level of statistical significance obtained in our epistasis screen.

Table 3.4. Interaction LOD scores for phenotypes significant at the genome-wide level ($P < 0.05$).

Phenotype	Chr A	Position (cM)	Chr B	Position (cM)	Lod.full[a]	Lod.fv1[b]	Lod.int[c]
SC	8	95	11	29	11.62	8.81	4.62
SL	2	60	9	5	7.45	5.87	4.25
TRS	9	97	11	6	13.00	8.02	6.49
LRD/Bz	7	57	8	81	9.26	7.75	6.97
SW	1	30	6	54	7.98	5.68	3.78
SW	6	54	9	87	8.49	6.20	3.77
SS	9	89	11	3	9.13	6.26	3.86

Two-way epistatic interactions for *S. lycopersicum / S. pimpinellifolium* RIL population across all 12 chromosomes.

[a] Lod.full is the LOD score of the full model with two loci and their interaction compared to the null model with no QTL.

[b] Lod.fv1 is the LOD score of the full model compared to the best single QTL model with one locus on either chromosome A or B.

[c] Lod.int is the LOD score of the interaction term which is found by comparing the full model with an interaction term, to the two QTL models with no interaction term.

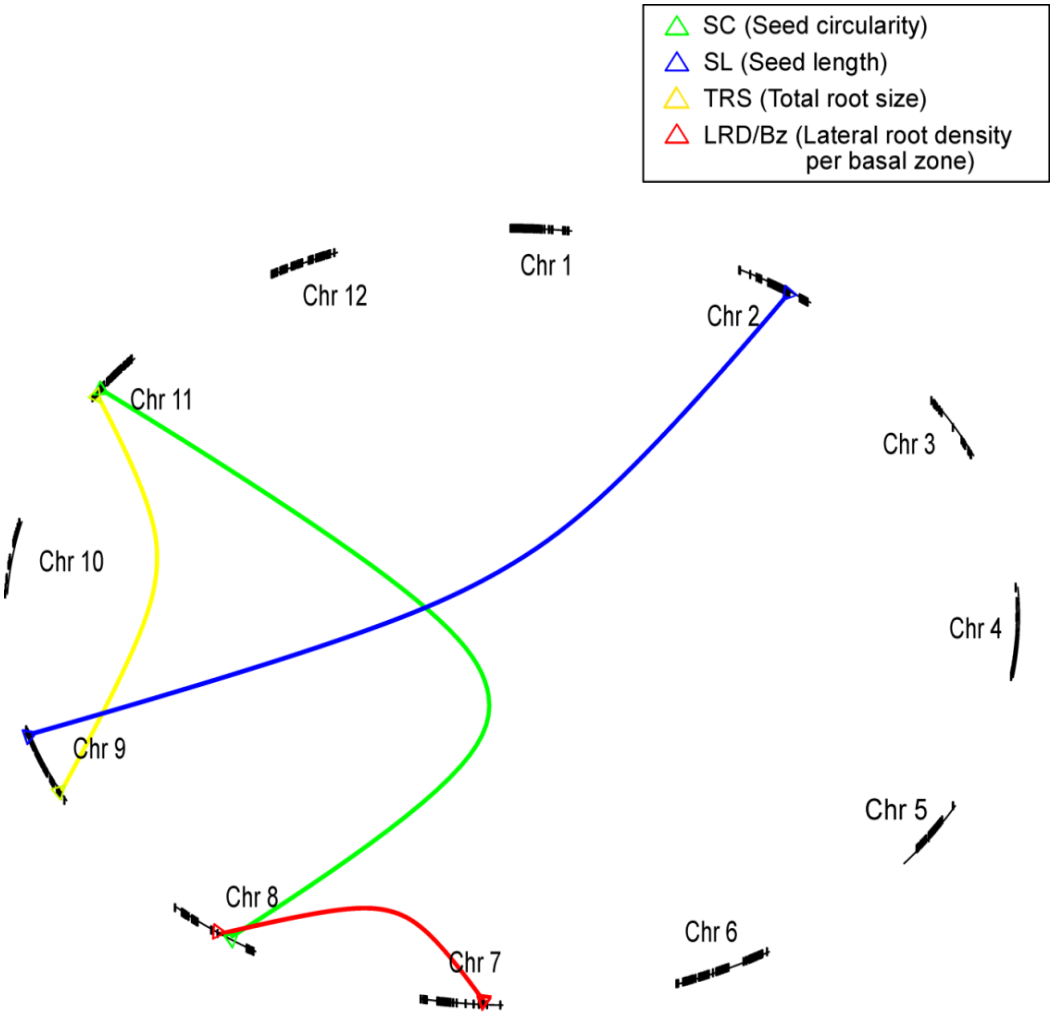


Figure 3.4. Epistatic interaction network of QTLs identified for seed and seedling quality traits. Graphical visualization of the epistatic interactions observed among different loci controlling seed and seedling quality phenotypes. The 12 chromosomes are represented as different circle

segments, and their sizes are proportional to the corresponding genetic sizes measured in cM. The color of the lines indicates the trait for which the epistatic interaction was observed (Arends *et al.* 2010).

Discussion

During our study we found considerable variation between the two parents for all the physiological parameters tested and an even higher variation was found in the RIL population, since transgression was observed for most of the traits. The phenotypic variation in the two parents, as well as in the RIL population and the resolution and size of this population was sufficient to find QTLs for seed and seedling quality, showing that this RIL population is a powerful tool for the study of the quantitative traits under study. We have utilized homogenous and strictly controlled plant growth conditions and seedling phenotype testing and this has contributed to the high genetic heritability that we observed for most of the traits. It furthermore indicates that the measured traits have a strong genetic regulation.

In a previous study (Kazmi et al., 2012) we analyzed 42 seed quality traits and identified 120 QTLs under optimal and stress conditions. Thus this population provides a valuable source for exploring the genes influencing complex phenotypes for seed quality as they allow isolation of the effect of a specific QTL from those of the entire genome and consequently enhance the statistical power to unravel quantitative seed quality phenotypes, controlling complex underlying mechanisms.

The seedling's ability for shoot penetration through the impeding soil of the seed bed is an essential attribute of vigor (Whalley et al., 1999). Rapid germination and subsequent seedling growth are, therefore, key phenotypes of vigorous seeds that are known to differ with genetic

background (Bettey et al., 2000). Thus, a vigorous seed must possess three key traits to establish seedlings across a wide range of environments: (1) the seed must germinate rapidly; (2) should have rapid initial downward growth; and (3) must have high potential for rapid upward shoot growth. Data obtained from fresh and dry root and shoot weights are good indicators for estimating the downward growth rate of root and upward growth rate of shoot, as well as predicting seed vigor (Bettey et al., 2000; Epstein, 2004; Fita et al., 2008).

Keeping in view the background and importance of seedling vigor through testing root and shoot growth of the seedling, we analyzed our RIL population for these traits and detected 10 QTLs for seedling growth on agar plates and 10 QTLs for growth of seedlings without nutrition. In addition, we identified 17 QTLs for seedling root architecture and 25 QTLs for seed dimension related traits. Most traits were enhanced by an allele of the *S. lycopersicum* parent, which displays vigorous seedling growth and high seed weight. However 27.5% of the detected QTLs had allelic effects enhanced by the *S. pimpinellifolium* parent, but these included QTLs for SC and LRD/Bz which indicates that small seeds have higher values for seed circularity and more lateral roots per basal zone in this population. Similar results were obtained in other tomato populations with the majority of the enhancing alleles for seed weight, fruit weight and total yield (Grandillo and Tanksley, 1996), and different botanical traits (DeVicente and Tanksley, 1993) coming from the *S. lycopersicum* parent. Our results are also supported by results in other crops in which QTLs were mainly affected by the positive allele of the parent with the heavy-weighted seed, for example in a study of the root architecture in melon (Fita et al., 2008). Besides the observed strong positive correlation between seed dimensions and seedling traits, we also found co-location of QTLs for these traits, as might be

expected from these results. Co-location of QTLs for different traits can be an indication that a locus has a pleiotropic effect on multiple traits, due to a common mechanistic basis or a dependency of traits (Clerkx et al., 2004). For example, a QTL on linkage group 9 is shared by five traits such as FrRt, DrRt, FrSh, DrSh and SW whereas the QTLs on linkage group 1 at marker position 69430752 are common between FrRtwn, SW and SS, respectively. In the present study most of the QTLs with major effect on all five seedling traits were identified on linkage groups 1, 6, 9 and 11. Most of these QTLs were co-locating with the QTLs for seed traits that we have identified in the current study and the QTLs identified in other studies of tomato seed weight (Tanksley et al., 1982; Weller et al., 1988; Goldman et al., 1995; Grandillo and Tanksley, 1996; Doganlar et al., 2000). These results are in agreement with those reported by Nieuwhof *et al.*, (1989), who tested 15 tomato genotypes with different seed size and 105 F1 obtained by di-allel crossing and found that genotypes with large seeds produced heavier seedlings than genotypes with small seeds. They also found a correlation between seed and seedling weight in the same range (R^2=0.8) as we have found in our study. The effect of seed weight on seedling growth may be due to the genetic variation in the amount of reserve food in the seeds and possible influenced by the maternal environment during seed development and maturation.

We found no significant correlations between seed size or seed weight and seed performance, such as rate and uniformity of germination or maximum germination percentage (Kazmi et al., 2012), as was also found in other species (Fenner, 1991). Thus, seed size is beneficial to the establishment of seedlings, but there appears to be no consistent link between seed size and germination characteristics.

Many selective factors affect seed size (Janzen, 1969; Harper et al., 1970; van der Pijl, 1972; Howe and Smallwood, 1982; Willson, 1983; Sorensen and Brodbeck, 1986; Fenner, 1991). The environment has great influence on seed size, with many factors that interact to affect the trait, such as high temperatures, short days, red light, drought and high nitrogen levels (Fenner, 1991). In tomato several studies have been carried out to identify QTLs for seed weight with seven different populations involving interspecific crosses between cultivated tomato and five wild tomato species (Tanksley et al., 1982; Weller et al., 1988; Goldman et al., 1995; Grandillo and Tanksley, 1996). The number of QTLs varied from 3 to 14 per study depending on the analytical method and the genetic populations used. In total 24 seed weight QTLs have been identified by different studies (Doganlar et al., 2000). Twelve seed weight QTLs were detected in only one species while 11 seed weight QTLs in two or more different species. One of the QTLs (*sw4.1*; Orsi and Tanksley, 2009) was common among all species and we found a QTL at the same position. In spite of the large number of QTLs identified for seed weight, no attention has been given in the previous studies to seed dimensions such as seed size and seed length. Although seed size, length and seed weight are closely related traits and are interdependent on each other, we measured differences in the total number of QTLs identified for seed weight (6 QTLs), seed size (4 QTLs) and seed length (3 QTLs) (Table 3.3), as well as in the individual and total explained variance of QTLs for seed weight (total exp. variance 60.9 %) , seed size (36.5 %) and seed length (33.3 %). The detected QTLs for seed size are co-locating with the seed weight QTLs, but 2 of the 3 seed length QTLs are found on different locations. This indicates that although a strong correlation can be expected between the different seed

dimension parameters, there are at least different loci influencing seed length as compared to seed size and weight.

A large number of QTLs for seed weight has also been identified in other crops. As an example, Teng *et al.*, (2008) found 94 QTLs for seed weight in soybean at different developmental stages. The identification of such a large number of QTLs for seed weight and the differences in the number and location of QTLs in different studies including the QTLs that we have detected for seed weight and size in our present study, illustrate that seed weight and seed dimensions are complex traits which are controlled by many genetic loci. In addition, the interaction of these loci with the environment may also affect the identification, location and number of QTLs as shown with the different numbers and positions of the seedling QTLs under two different environmental conditions (Table 3.3).

There is experimental evidence that larger seeds are better able to establish or survive as seedling in a variety of environments, including nutrient shortage (Lee and Fenner, 1989; Jurado and Westoby, 1992). This corroborates our observation of a greater correlation between seed weight and seedling vigor under nutrient-deprived condition than on MS medium with nutrients (Figure3.2). In general the shoot and root weights of the two parents as well as those in the RIL population were significantly lower under the nutrient-deprived conditions than those on vertical agar plates with MS nutrition. These results are in agreement with those reported by Nieuwhof *et al.*, (1989), who observed significant correlation between tomato seed size and seedling mass under nutrient-deprived conditions. We also observed some differences in the identification of QTLs between the two experiments. In general we identified higher numbers of QTLs with higher explained variance for three seedling traits (FrRtwn, DrRtwn FrShwn) in nutrient-deprived conditions (Table 3.3, Figure 3.3). For the

nutrient deprived conditions, 9 out of 10 QTLs are overlapping with SW/SS QTLs, while for the growth of seedling with nutrients, 5 out of 7 seedling trait QTLs and 2 out of 3 HypL QTLs overlap with SW/SS QTLs. Although most seedling QTLs overlapped with seed dimension QTLs, we found some exceptions. A QTL for FrRt and HypL was found on chromosome 10 with explained variances from 8.5 and 9.3% respectively and another QTL on chromosome 12 for FrRt with an explained variance of 8.2%. Additionally a QTL for FrRtwn was found on chromosome 7 with an explained variance of 6.5%. The detection of these loci suggests the possibility for breeding for seedling vigor independent of seed size.

Genotypes x environment interactions are very important for the expression of QTLs. In the present study identification of different QTLs in both of the environments indicates that some QTLs seem to be sensitive to the environment, but a substantial proportion of QTLs was found in both experiments. Especially the QTLs with higher LOD scores for all the traits could readily be detected in both environments. Therefore, the present study tends to support the general conclusion made by Tanksley, 1993, who concluded that a substantial proportion of QTLs affecting a trait can be identified under different environments, especially QTLs that have major effects.

Root systems execute the crucial task of providing water, nutrients and physical support to the plant. The length of the primary/main root and the number of the lateral roots determine the architecture of the root system. This root system in turn, plays a major role in determining whether a plant will succeed in a particular environment (Malamy and Benfey, 1997). A fast-growing and improved deep root system will improve competitiveness with weeds during the initial stage of seedling growth. Furthermore it will also be more efficient in the acquisition of nutrients and

uptake of water from lower layers of soil during low-nutrient- and low-moisture conditions. In soil or media with a patchy nutrient distribution, lateral roots preferentially proliferate in the nutrient-rich zone (Robinson, 1994; Zhang et al., 1999) and thereby play an important role in the uniform utilization of nutrients from the soil. There are some studies which, in addition to its effect on the upward growth of seedlings, also demonstrate a correlation between seed traits (seed weight, -size and -vigor) on the initial downward growth of the root system (Baker, 1972; Jurado and Westoby, 1992). Finch-Savage *et al.*, (2010) found strong effects of seed vigor in *Brassica oleracea* on the initial downward growth of seedlings and fine mapped QTLs for rapid initial growth of root which also co-located with seed weight.

As the underground parts of plants are difficult to quantify, studies on roots are lagging behind those of shoots (Epstein, 2004). In the case of tomato no relevant information is available on root growth related traits nor has any proper study on seedling growth been published and, therefore, to the best of our knowledge, this is the first genetic analysis of seedling traits in tomato. Our results on root architecture tend to support the argument that larger food reserves in large-sized seed help in establishing an extensive root system. We observed that the heavy-weighted seed parent *S. lycopersicum* has a very strong root system with two times faster downwards growth (MRL=8.54 cm) and two times bigger total root size (TRS =13.99 cm) than the light-weighted seed parent *S. pimpinellifolium* with slow downward growth (MRL=4.61 cm) and small total root size (TRS=6.36 cm). These results are in agreement with the phenotypic values of fresh and dry root weights of the two parents. In total we identified 5 QTLs for MRL and 6 QTLs for TRS. For three major QTLs for MRL and for all the TRS QTLs, the positive alleles are derived from the *S. lycopersicum*

parent (Table 3.3 and Figure 3.3). In both of these cases, the major effect QTLs were also co-locating with SW and SS QTLs on linkage groups 9 and 11. On the other hand, the QTLs for LRn and LRD/Bz had 50% of the positive alleles from both parents with some major QTLs from the *S. lycopersicum* parent and these major QTLs were also co-locating with the seed size QTLs. The LRD/Bz value is relatively high for *S. pimpinellifolium*. This result illustrates that *S. pimpinellifolium* has a short branched zone with a high density of lateral roots, while *S. lycopersicum* has a longer branched zone with a lower density of lateral roots.

The co-location of QTLs for MRL, TRS, LRn, LRD/Bz and seed dimension traits with the positive additive effects from the same parent and the correlation of the phenotypic values for these traits, indicates that root and seed traits may be genetically interlinked traits and may be under the control of common genetic mechanisms.

For all the co-locations found in this study, it is not known whether it is a common allele controlling all the traits or whether it is a cluster of different alleles for different traits located closely together. Classical quantitative genetics assumes that trait correlation can be due to the effect of pleiotropy or due to the tight linkage of genes. For pleiotropic effects, one can expect not only the same location of QTLs for related traits, but also the same direction of their allelic effects. If close linkage of genes was the major reason, the directions of the genetic effects of the QTLs for different traits may be different, although coincidence of QTL locations can still be expected. The fact that most favorable alleles for the QTLs described in this study have been derived from the *S. lycopersicum* parent might suggest that pleiotropy rather than close linkage of different alleles is the major reason for correlation of the measured traits. In general, we found a high correlation between seed and seedling traits, but although we

found co-localization of some RSA QTLs with seed dimension QTLs, the overall correlation between these traits was low. Eight out of the 17 RSA QTLs do not co-locate with seed dimension QTLs. These include major QTLs for LRn on chromosome 5 and for MRL on chromosome 7 with explained variances of 13 and 6.5% respectively and minor QTLs on chromosome 1 for MRL and TRS with explained variances of 6.1 and 5.6% respectively and on chromosome 3 and 10 for TRS explaining respectively 5.9 and 4.8 % variance. These RSA QTLs together with the previous mentioned seeds size independent seedling weight QTLs indicate that in addition to seed size there are other mechanisms involved in controlling seedling establishment under different environmental conditions.

In conclusion, the strong co-location of QTLs among different seed and seedling traits with generally the same genetic direction of the QTLs and the correlation in the phenotypic values of these traits, indicate a strong correlation among seed- and seedling vigor and seed size and weight appear to have a strong effect on the initial downward growth of the main root and upward growth of the shoot. This positive effect of heavy seed could be due to common genetic mechanisms controlling these traits and also to the high quantity of reserve food in larger seeds as compared to small seeds.

Apart from the correlation between seed and seedling traits we also tested the correlation between seed weight and seed performance in a previous analysis (Kazmi et al., 2012), but found no significant correlation between seed weight and germination rate (t_{50}), uniformity (U_{7525}) and final germination percentage (G_{max}%). Thus, increased seed size seems a benefit for seedling establishment, but a consistent link between seed size and germination characteristics is not obvious. In tomato it has been reported that inheritance of time to germination was closely related to seed

size, with the smaller seeds germinating earlier (Whittington, 1973). However, our data show that this is not the case for the here studied population. Furthermore we have also shown that germination performance and seed size are controlled by different independent genetic loci (Kazmi et al., 2012).

The mapping of QTLs associated with key seed- and seedling-vigor traits in tomato could open up various opportunities to improve efficiency of plant breeding and selection for lines with improved seed vigor and, hence, seedling and crop establishment. Molecular markers linked to the QTLs may be utilized in marker-assisted selection, providing a rapid method to select for specific genotypes without the need to extensively assess phenotypes at all stages in the breeding program. Furthermore, we will follow up the defined QTLs with fine-mapping and improvement of candidate-gene selection by the use of a genetical genomics set-up and thereby elucidate the molecular mechanisms that control seed- and seedling-vigor (Joosen et al., 2009; Ligterink et al., 2012).

Acknowledgments

This work was supported by the Technology Foundation STW (R.K., L.W., W.L.) and by the Higher Education Commission, Pakistan (N.K.).

References:

Arbuckle J (1999) Amos 4.0 user's guide. Marketing Department, SPSS Inc.

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

Armengaud P, Zambaux K, Hills A, Sulpice R, Pattison R, Blatt M, Amtmann A (2009) EZ Rhizo: integrated software for the fast and accurate measurement of root system architecture. The Plant Journal **57:** 945-956

Baker H (1972) Seed weight in relation to environmental conditions in California. Ecology **53:** 997-1010

Bettey M, Finch-Savage W, King G, Lynn J (2000) Quantitative genetic analysis of seed vigour and pre-emergence seedling growth traits in Brassica oleracea. New Phytologist **148:** 277-286

Bleasdale J (1967) The relationship between the weight of a plant part and total weight as affected by plant density. Journal of Horticultural Science **42:** 51-58

Broman KW, Wu H, Sen S, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. Bioinformatics **19:** 889-890

Cannon O, Gatherum D, Miles W (1973) Heritability of low temperature seed germination in tomato. HortScience **8:** 404-405

Castro J, Hodar J, Gomez J (2006) Seed size. Handbook of seed science and technology**:** 397

Chambers J (1995) Relationships between seed fates and seedling establishment in an alpine ecosystem. Ecology **76:** 2124-2133

Clerkx E, El-Lithy M, Vierling E, Ruys G, Blankestijn-De Vries H, Groot S, Vreugdenhil D, Koornneef M (2004) Analysis of natural allelic variation of Arabidopsis seed germination and seed longevity traits between the accessions Landsberg erecta and Shakdara, using a new recombinant inbred line population. Plant Physiology **135:** 432

Coolbear P (1995) Mechanisms of seed deterioration. *In* A Basra, ed, Seed Quality: Basic Mechanisms and Agricultural Implications. Food Products Press, New York pp 223-277

Cornelissen J (1999) A triangular relationship between leaf size and seed size among woody species: allometry, ontogeny, ecology and taxonomy. Oecologia **118:** 248-255

DeVicente M, Tanksley S (1993) QTL analysis of transgressive segregation in an interspecific tomato cross. Genetics **134:** 585

Dickson M (1980) Genetic aspects of seed quality. Horticultural Science **15:** 771-774

Doerge RW, Churchill GA (1996) Permutation tests for multiple loci affecting a quantitative character. Genetics **142:** 285-294

Doganlar S, Frary A, Ku H, Tanksley S (2002) Mapping quantitative trait loci in inbred backcross lines of Lycopersicon pimpinellifolium (LA1589). Genome **45:** 1189-1202

Doganlar S, Frary A, Tanksley S (2000) The genetic basis of seed-weight variation: tomato as a model system. Theoretical and Applied Genetics **100:** 1267-1273

Dolan R (1984) The effect of seed size and maternal source on individual size in a population of Ludwigia leptocarpa (Onagraceae). American Journal of Botany**:** 1302-1307

Ellis R (1992) Seed and seedling vigour in relation to crop growth and yield. Plant growth regulation **11:** 249-255

Epstein E (2004) Plant biologists need to get back to their roots. Nature **430:** 829

Fenner M (1991) The effects of the parent environment on seed germinability. Seed Science Research **1:** 75-84

Finch-Savage W (1995) Influence of seed quality on crop establishment, growth and yield. Seed Quality: Basic Mechanisms and Agricultural Implications**:** 361-384

Finch-Savage W, Clay H, Lynn J, Morris K (2010) Towards a genetic understanding of seed vigour in small-seeded crops using natural variation in Brassica oleracea. Plant Science

Fita A, Pico B, Monforte A, Nuez F (2008) Genetics of Root System Architecture Using Near-isogenic Lines of Melon. Journal of the American Society for Horticultural Science **133:** 448-458

Foolad M, Chen F, Lin G (1998) RFLP mapping of QTLs conferring cold tolerance during seed germination in an interspecific cross of tomato. Molecular Breeding **4:** 519-529

Foolad M, Stoltz T, Dervinis C, Rodriguez R, Jones R (1997) Mapping QTLs conferring salt tolerance during germination in tomato by selective genotyping. Molecular Breeding **3:** 269-277

Foolad M, Subbiah P, Zhang L (2007) Common QTL affect the rate of tomato seed germination under different stress and nonstress conditions. International journal of plant genomics **2007:** 97386

Goldman I, Paran I, Zamir D (1995) Quantitative trait locus analysis of a recombinant inbred line population derived from a Lycopersicon esculentum x Lycopersicon cheesmanii cross. Theoretical and Applied Genetics **90:** 925-932

Grandillo S, Tanksley S (1996) QTL analysis of horticultural traits differentiating the cultivated tomato from the closely related species Lycopersicon pimpinellifolium. Theoretical and Applied Genetics **92:** 935-951

Greene D, Johnson E (1998) Seed mass and early survivorship of tree species in upland clearings and shelterwoods. Canadian Journal of Forest Research **28:** 1307-1316

Harper J, Lovell P, Moore K (1970) The shapes and sizes of seeds. Annual Review of Ecology and Systematics **1:** 327-356

Hilhorst H, Koornneef M (2007) Dormancy in Plants. *In* Encyclopedia of life sciences Wiley, Cichester, pp 1-4

Hilhorst H, Toorop P (1997) Review on dormancy, germinability, and germination in crop and weed seeds. Advances in Agronomy **61:** 111-165

Howe H, Smallwood J (1982) Ecology of seed dispersal. Annual Review of Ecology and Systematics **13:** 201-228

Janzen D (1969) Seed-eaters versus seed size, number, toxicity and dispersal. Evolution**:** 1-27

Jones R (1986) High salt tolerance potential in Lycopersicon species during germination. Euphytica **35:** 575-582

Joosen R, Kodde J, Willems L, Ligterink W, van der Plas L, Hilhorst H (2010) germinator: a software package for high-throughput scoring and curve fitting of Arabidopsis seed germination. The Plant Journal **62:** 148-159

Joosen RV, Ligterink W, Hilhorst HW, Keurentjes JJ (2009) Advances in genetical genomics of plants. Current Genomics **10:** 540-549

Jurado E, Westoby M (1992) Seedling growth in relation to seed size among species of arid australia. Journal of Ecology **80:** 407-416

Kazmi RH, Khan N, Willems LA, AW VANH, Ligterink W, Hilhorst HW (2012) Complex genetics controls natural variation among seed quality phenotypes in a recombinant inbred population of an interspecific cross between Solanum lycopersicum x Solanum pimpinellifolium. Plant, Cell & Environment **35:** 929-951

Keurentjes JJ, Bentsink L, Alonso-Blanco C, Hanhart CJ, Blankestijn-De Vries H, Effgen S, Vreugdenhil D, Koornneef M (2007) Development of a near-isogenic line population of Arabidopsis thaliana and comparison of mapping power with a recombinant inbred line population. Genetics **175:** 891-905

Koller DL, Liu LX, Alam I, Sun QW, Econs MJ, Foroud T, Turner CH (2009) Epistasis between QTLs for bone density variation in Copenhagen x dark agouti F2 rats. Mammalian Genome **20:** 180-186

Koornneef M, Bentsink L, Hilhorst H (2002) Seed dormancy and germination. Current opinion in plant biology **5:** 33-36

Lee W, Fenner M (1989) Mineral nutrient allocation in seeds and shoots of twelve Chionochloa species in relation to soil fertility. The Journal of Ecology **77:** 704-716

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

Maas E (1986) Salt tolerance of plants. Applied Agricultural Research **1:** 12-25

Malamy JE, Benfey PN (1997) Down and out in Arabidopsis: The formation of lateral roots. Trends in Plant Science **2:** 390-396

Marshall D (1986) Effect of seed size on seedling success in three species of Sesbania (Fabaceae). American Journal of Botany **73:** 457-464

McDonald M (1998) Seed quality assessment. Seed Science Research **8:** 265-276

Moegenburg S (1996) Sabal palmetto seed size: causes of variation, choices of predators, and consequences for seedlings. Oecologia **106:** 539-543

Morse D, Schmitt J (1985) Propagule size, dispersal ability, and seedling performance in Asclepias syriaca. Oecologia **67:** 372-379

Nieuwhof M, Garretsen F, Oeveren J (1989) Maternal and genetic effects on seed weight of tomato, and effects of seed weight on growth of genotypes of tomato (Lycopersicon esculentum Mill.). Plant Breeding **102:** 248-254

Orsi CH, Tanksley SD (2009) Natural Variation in an ABC Transporter Gene Associated with Seed Size Evolution in Tomato Species. PLoS Genetics **5:** e1000347

Rick C (1958) The role of natural hybridization in the derivation of cultivated tomatoes of western South America. Economic Botany **12:** 346-367

Rick C (1973) Potential genetic resources in tomato species: clues from observations in native habitats. Basic Life Sciences **2:** 255

Rick C (1982) The potential of exotic germplasm for tomato improvement. *In* I Vasil, ed, Plant Improvement and Somatic Cell Genetics. Academic Press, New York, pp 1-27

Rick C, Fobes J, Holle M (1977) Genetic variation inLycopersicon pimpinellifolium: Evidence of evolutionary change in mating systems. Plant Systematics and Evolution **127:** 139-170

Robinson D (1994) The responses of plants to non-uniform supplies of nutrients. New Phytologist **127:** 635-674

Scott S, Jones R (1982) Low temperature seed germination of Lycopersicon species evaluated by survival analysis. Euphytica **31:** 869-883

Seiwa K, Kikuzawa K (1991) Phenology of tree seedlings in relation to seed size. Canadian Journal of Botany **69:** 532-538

Seiwa K, Kikuzawa K (1996) Importance of seed size for the establishment of seedlings of five deciduous broad-leaved tree species. Plant Ecology **123:** 51-64

Sorensen K, Brodbeck U (1986) A sensitive protein assay method using micro-titer plates. Cellular and Molecular Life Sciences **42:** 161-162

Stanton M (1984) Seed variation in wild radish: effect of seed size on components of seedling and adult fitness. Ecology **65:** 1105-1112.

Stebbins G (1976) Seed and seedling ecology in annual legumes. I. A comparison of seed size and seedling development in some annual species. Oecologia Plantarum **11:** 321-331

Stubbe H (1960) Mutanten der WildtomateLycopersicon pimpinellifolium (Jusl.) Mill. I. Genetic Resources and Crop Evolution **8:** 110-137

Stubbe H (1965) Mutanten der WildtomateLycopersicon pimpinellifolium (Jusl.) Mill. III. Genetic Resources and Crop Evolution **13:** 517-544

Tanksley S (1993) Mapping polygenes. Annual Review of Genetics **27:** 205-233

Tanksley S, Medina-Filho H, Rick C (1982) Use of naturally-occurring enzyme variation to detect and map genes controlling quantitative traits in an interspecific backcross of tomato. Heredity **49:** 11–25

Teng W, Han Y, Du Y, Sun D, Zhang Z, Qiu L, Sun G, Li W (2008) QTL analyses of seed weight during the development of soybean (Glycine max L. Merr.). Heredity **102:** 372-380

Tripathi R, Khan M (1990) Effects of seed weight and microsite characteristics on germination and seedling fitness in two species of Quercus in a subtropical wet hill forest. Oikos**:** 289-296

van der Pijl L (1972) Principles of dispersal in higher plants. Springer-Verlag, Heidelberg

Van Ooijen JW, Maliepaard C (2003) MapQTL®, Version 5.0: Software for the Calculation of QTL Positions on Genetic Maps. *In*. Institute of Plant Genetics, Polish Academy of Sciences, p 305

Voorrips R, Verkerke W, Finkers R, Jongerius R, Kanne J (2000) Inheritance of taste components in tomato. Acta Physiologiae Plantarum **22:** 259-261

Weller J, Soller M, Brody T (1988) Linkage analysis of quantitative traits in an interspecific cross of tomato (Lycopersicon esculentum x Lycopersicon pimpinellifolium) by means of genetic markers. Genetics **118:** 329

Whalley W, Finch-Savage W, Cope R, Rowse H, Bird N (1999) The response of carrot (Daucus carota L.) and onion (Allium cepa L.) seedlings to mechanical impedance and water stress at sub-optimal temperatures. Plant, Cell & Environment **22:** 229-242

Whittington W (1973) Genetic regulation of germination. Seed ecology: proceedings**:** 5

Willson M (1983) Plant reproductive ecology. New York

Winn A (1988) Ecological and evolutionary consequences of seed size in Prunella vulgaris. Ecology **69:** 1537-1544

Wolf S, Yakir D, Stevens M, Rudich J (1986) Cold temperature tolerance of wild tomato species. Journal of the American Society for Horticultural Science **111:** 960-964

Wood D, Morris W (1990) Ecological constraints to seedling establishment on the pumice plains, Mount St. Helens, Washington. American Journal of Botany **77:** 1411-1418

Wright I, Westoby M (1999) Differences in seedling growth behaviour among species: trait correlations across species, and trait shifts along nutrient compared to rainfall gradients. Journal of Ecology **87:** 85-97

Wudiri B, Henderson D (1985) Effects of water stress on flowering and fruit set in processing-tomatoes. Scientia Horticulturae **27:** 189-198

Wulff R (1986) Seed size variation in Desmodium paniculatum: II. Effects on seedling growth and physiological performance. The Journal of Ecology **74:** 99-114

Zhang HM, Jennings A, Barlow PW, Forde BG (1999) Dual pathways for regulation of root branching by nitrate. Proceedings of the National Academy of Sciences of the United States of America **96:** 6529-6534

Zhang J, Maun M (1991) Establishment and growth of Panicum virgatum L. seedlings on a Lake Erie sand dune. Bulletin of the Torrey Botanical Club **118:** 141-153.

Chapter 4

METABOLOMICS - THE LINK BETWEEN GENOTYPE, METABOLOME AND PHENOTYPE FOR SEED QUALITY TRAITS IN TOMATO

Kazmi RH, Willems LAJ, Ligterink W, Hilhorst HWM

In preparation

Abstract

Genomic approaches have accelerated the study of the quantitative genetics that underlie phenotypic variation. We have utilized gas chromatography-time-of-flight/mass spectrometry (GC-TOF-MS) metabolite profiling to characterize tomato seeds during dry and imbibed stages using a recombinant inbred line (RIL) population of *Solanum lycopersicum* x *Solanum pimpinellifolium*. In this article we describe, for the first time, the use of a generalized genetical genomics (GGG) model to study metabolite changes in tomato seeds incorporating genetics as well as environmental effects. The GGG design was used to map genetic quantitative trait loci (G QTLs) and environmental changes by genetic-by-environment interactions (G x E QTLs). A significant canonical correlation was found between metabolites and seed quality traits, revealing a close link between seed quality phenotypes and a specific combination of metabolites. The empirical finding in this study provides a new understanding of the high predictive power of the metabolic composition for seed quality phenotypes. The topologies of metabolic correlations that we have identified may contribute

to enhancing our understanding of the role of highly connected metabolites in these metabolic networks and their specific functions in seed-developmental biochemical pathways. Densely connected metabolites were extracted using graph clustering from correlation networks, and the clusters were evaluated by biochemical-pathway enrichment analysis. The evidence from this study suggests that the number of significant correlations varied among individual metabolites and that the obtained clusters were significantly enriched for metabolites involved in specific biochemical pathways. Thus, the methods described here have the potential to reveal regulatory networks that contribute to our understanding of the complex nature of seed quality.

Introduction

Genomic approaches have accelerated the study of the quantitative genetics that underlie phenotypic variation. The mutualistic relationship between metabolomics and genetics goes back to Mendel's reliance upon metabolic phenotypes (anthocyanins and starch) to develop his basic genetic theory (Kliebenstein, 2009). The understanding of DNA structure and metabolism was further enhanced as genetics has played an equally important role in the reconstruction of biochemical pathways eventually shaping our current understanding of gene regulation. The combination of metabolomics and genetics has provided powerful insights into the origin and maintenance of natural variation (Keurentjes et al., 2006). Given their huge diversity, metabolites can be associated with specific genetic markers, mRNA transcripts, and enzyme activities, allowing a linkage between variation from genetic to biochemical levels that is more complex for less-defined or more pleiotropic phenotypes, such as seed quality (Koornneef et al., 2004; Keurentjes et al., 2008; Keurentjes and Sulpice, 2009).

Recent advances in genome sequencing and high-throughput phenotyping technologies have enabled plant biologists to study the quantitative genetics and association of genetic variation to phenotypic variation on a comprehensive large scale analysis basis (Keurentjes, 2009; Chan et al., 2010). The combination of metabolomics with quantitative genetics is at the heart of our understanding of biochemical phenotypes. Correspondingly, the fitness consequences of these metabolic changes are an important component in the determination of the genetic architecture of species, making metabolomics unique in the quest for system-wide coverage of all metabolites (Phillips, 2008; Kliebenstein, 2009). The ultimate aim of metabolomics is to quantify the level of all intermediates of metabolism (Buscher et al., 2009). Research has consistently shown that quantitative metabolomics data can directly be mapped onto the metabolic network, ultimately opening the door for identification of metabolic reactions, networks and biochemical pathways (Rowe et al., 2008; Keurentjes, 2009; Sulpice et al., 2010). The combination of disciplines has broadened our fundamental apprehension of how metabolites can provide biotic and abiotic stress tolerance in crops like *Zea mays* (Byrne et al., 1998), *Brassica* (Mithen and Magrath, 1992; Magrath et al., 1993), *Lotus japonicus* (Sanchez et al., 2011), tomato (Rajasekaran et al., 2001) and Arabidopsis (Kaplan et al., 2004; Meyer et al., 2007). Several studies have demonstrated the use of metabolic QTLs (mQTLs) in integration of different levels of genomic information (sequence, transcript, and protein) to understand plant and seed phenotypes better, improve crop breeding and obtain ecological inference about the corresponding selective pressure acting on these QTLs (Schauer et al., 2006; Lisec et al., 2008; Schauer et al., 2008; Lisec et al., 2009).

Metabolites can be categorized as primary or secondary although it is difficult to make a clear distinction as interactions between the two categories are ominously present (Carrari et al., 2006). There is an unambiguous relationship between primary metabolism and plant growth and development (Lisec et al., 2008). Contrary to secondary metabolism, major perturbation of the networks of the primary metabolism has strong detrimental effects on plant performance. This is supported by numerous transgenic approaches showing that perturbation of pathways of secondary metabolites, such as carotenoids or flavonoids, can be achieved without any major pleiotropic effects concerning growth and development. In contrast, when attempting to modify primary metabolism, such as sucrose biosynthesis or the tricarboxylic acid (TCA) cycle, a negative impact on growth and development is usually found to be a consequence, at the whole-plant level (Trethewey et al., 1998). What we know about primary metabolites is largely based upon empirical studies that show that they function in central carbohydrate metabolism (Price et al., 2008), whereas secondary metabolites are often connected to cell signaling, interspecies communication and responses to biotic and abiotic stress (Wobus and Weber, 1999; Garg et al., 2002; Scheible et al., 2004).

Variations in plant growth, as well as in seed and metabolic traits, have been detected for a series of natural accessions and recombinant inbred lines (Schauer et al., 2006; Meyer et al., 2007; Prinzenberg et al., 2010; Skogerson et al., 2010; Toubiana et al., 2012). Although only weak relationships have been suggested between growth and the levels of individual metabolites (Meyer et al., 2007), highly significant links between biomass and a specific combination of metabolites have been demonstrated (Lisec et al., 2008; Prinzenberg et al., 2010).

Metabolite profiling in Arabidopsis during seed development (Fait et

al., 2006) identified major metabolic abundance switches associated with successive developmental stages. Although certain alterations that impair cellular structures and metabolism have been implicated in seed deterioration, the molecular and biochemical basis of seed quality is not well understood. Despite extensive research on metabolic profiling for tomato fruit quality, no single study exists which adequately covers the relationship between metabolite profiles and the modular operation of biochemical networks relevant to tomato seeds (Schauer et al., 2006; Schauer et al., 2008; Toubiana et al., 2012).

Recent developments in "omics" technologies have heightened the integration of different data sets. A genetical genomics approach brings together traditional QTL mapping with gene expression and metabolic profiling studies for a better understanding of the genetic mechanisms influencing complex traits (Jansen and Nap, 2001). This is a useful methodology in studying molecular perturbation in biological systems. Several studies have used this approach, focusing on natural variation (Keurentjes et al., 2006), the connection between metabolism and yield-associated traits or biomass (Schauer et al., 2006; Meyer et al., 2007), and the identification of metabolic quantitative trait loci (mQTL) (Lisec et al., 2008). Most studies using genetical genomics have only been carried out in *A. thaliana* mainly due to the availability of high quality mapping populations and the commercially available genome-wide micro-arrays where several studies in various RIL populations have indicated extensive genetic regulation of gene expression (Keurentjes et al., 2007; West et al., 2007). However, little attention has been paid to tomato in particular with respect to seed quality evaluation. In addition to molecular networks, the genetic perturbations of biological systems also depend upon environmental conditions and, thus, a comprehensive understanding of

biological systems requires studying them across multiple environments. We applied a generalized genetical genomics (GGG) approach for metabolic profiling using GC-TOF-MS on 83 recombinant inbred lines (RILs) of tomato to describe the genetic regulation of variation in the tomato seed metabolome. This new GGG model may prove to be useful in tomato seeds and allows the investigation of the mechanisms that contribute to complex variations in the tomato seed metabolome during germination.

Germination efficiency is affected by reserve accumulation during seed development or their mobilization during seed germination as well as several unknown factors (Fait et al., 2006). To elucidate the nature of such factors, we analyzed the metabolite content of tomato seeds at two developmental stages: dry and 6h-imbibed. Metabolic fluxes are arrested in the dry seed; however, upon imbibition the dry seed rapidly resumes metabolic activity (Bewley and Black, 1994). We chose the 6h stage for optimum synchronization of seed germination as full rehydration of dry seeds typically completes in less than 2 hours, and assuming that many metabolic processes will have started after 6 hours of imbibition. Thus, it was essential to have intelligent selection of time point for imbibed seeds as germination extends from the onset of imbibition in an environment meeting the normal physiological requirements for germination to the inception of cell division and elongation (Bewley and Black, 1994). The application of a GGG model, which is a systems genetics approach, provides a broad overview of changes in primary metabolic processes that occur during dry and imbibed tomato seed developmental stages. In particular, it takes into account genetics and chosen environmental perturbations (different seed developmental stages, i.e. dry and imbibed seeds) in combination with the analysis of the genetic variation present in F_8 RILs, to study the multiple environments and to identify genotype-by-

environment interactions. Thus, the present approach reveals, for the first time, the plasticity of molecular networks in tomato for seed quality traits and forms a crucial step toward understanding different influences of genetic and developmental responses in tomato seeds. The present study attempts to link seed traits to metabolic signatures. Furthermore, it supports previous findings in other crops and provides additional evidence that relationships between a seed trait and a single metabolite is generally absent but that strong canonical correlation with a specific combination of metabolites illustrates the complexity of such quantitative traits (Meyer et al., 2007).

Materials and Methods

Growth conditions and seed collection

The *S. lycopersicum* × *S. pimpinellifolium* RIL population was grown twice under controlled conditions in the greenhouse facilities at Wageningen University, in the Netherlands. The day and night temperatures were maintained at 25 °C and 15 °C, respectively, with 16h light and 8h dark (long-day conditions). All the RILs were uniformly supplied with the basic dose of fertilizer. Seeds were extracted from healthy fruits and treated with 1% hydrochloric acid (HCL) to remove the large pieces of the pulp that were sticking onto the seeds. The solution of tomato seed extract with diluted hydrochloric acid was passed through a fine mesh sieve and washed with water to remove the remaining parts of the pulp and remnants of the hydrochloric acid. The seeds were processed and disinfected by soaking in a solution of trisodium phosphate ($Na_3PO_4.12H_2O$) for 1h. Finally, the seeds were dried on clean filter paper at room temperature and were brushed to remove impurities with a seed brusher (Seed Processing Holland BV, Enkhuizen, The Netherlands, http://www.seedprocessing.nl). The cleaned seeds were dried for 3d at

20°C and were stored in a cool, dry storage room (13 °C and 30% RH) in paper bags.

Extraction, Derivatization, and Analysis of Seed Metabolites Using GC-TOF-MS

The extraction method is modified from the method previously described by (Roessner et al., 2000). Approximately 30mg seeds were homogenized in 2 ml tubes with 2 iron balls (2.5mm), pre-cooled in liquid nitrogen. For the homogenization the micro-dismembrator (Sartorius) is used at 1500 rpm. 700µl methanol/chloroform (4:3) was added together with the standard (0.2mg/ml ribitol) and mixed thoroughly. After 10 minutes sonication, 200µl MQ was added to the mixture followed by vortexing and centrifuging (5 mins 13,500rpm). Methanol phase was collected in a glass vial. 500µl methanol/chloroform was added to the remaining organic phase and kept on ice for 10 minutes. 200µl MQ was added followed by vortexing and centrifuging (5 mins 13,500rpm). Again, the methanol phase was collected and mixed with the other collected phase. 100µl was dried overnight in a speedvac (35 °C Savant SPD121). The GC-TOF-MS method was previously described by (Carreno-Quintero et al., 2012) 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 the mass signals that were present in less than 3 RILs were discarded. Out of all the mass signals, centrotypes are formed using the MSclust program (Tikunov et al., 2011). This resulted in 160 unique centrotypes

(representative masses). The mass spectra of these centrotypes were used for identification by matching to 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 with retention indices calculated by using a 3rd order polynomial function (Strehmel et al., 2008).

Statistical Analysis of GC-TOF-MS Data

Metabolomic data were log10 transformed and then statistically analyzed using the rank product method (Breitling et al., 2004) to identify differentially changed metabolites with the Bioconductor 'RankProd' package. Significantly changed metabolites showed a false discovery rate (FDR) < 0.05. The FDR value in the rank product was obtained with 1,000 random permutations. Principal component analysis was performed on the data sets obtained from metabolite profiling with the Bioconductor "pcaMethods" package (Stacklies et al., 2007). The data were log transformed and normalized to the median of the entire sample set for each metabolite before analysis. This transformation reduces the influence of outliers. Heat map presentation and clustering were performed with Spearman correlation coefficient matrices R-packages "MASS", "Hmisc" "VGAM" and their presentation as heat maps using R-packages "gplots" and "graphics" were used; also ANOVA was performed using R statistics (http://www.r-project.org/).

QTL Analysis

Data was pre-processed 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 ($y_i=\beta_0+\beta_1 g_i+\varepsilon_i$) on the two conditions separately. This was followed by a combined mapping allowing for a developmental co-variate and interaction term between the genetic marker and the developmental stage ($y_i= \beta_0+\beta_1 e_i+\beta_2 g_i+\beta_3 e_i{:}g_i+\varepsilon_i$). P-values from all mappings were transformed into LOD scores by taking the –log10. Additionally, raw and normalized effects were calculated for each individual environment. Normalized effects were calculated by dividing the difference between the maximum and minimum values for that trait by the mean effect at the marker. LOD significance was determined using permutations for the combined mapping of the two environments: an LOD score of 3 was found to be significant (Breitling et al., 2008).

Integrated Analysis of Phenotypic and Metabolite Data

The relationship between seed quality phenotypes and metabolite profile was analyzed using Spearman correlation between the seed quality phenotypes and relative abundances of all metabolites, and by a more complex multiplicative model (Meyer et al., 2007). Missing values in the metabolite matrix were imputed with a self-organizing map (SOM) algorithm using R package "SeqKnn".

Canonical correlation analysis (CCA)

Canonical correlation analysis calculates the highest possible correlation between linear combinations of the columns from two matrices with the same number of rows. The R function "cancor" was used to calculate the canonical correlation between metabolites and seed quality phenotypes. For cross validation a partial least square (PLS) regression was performed. To carry out the procedure the "pls" R-package implementing partial least squares regression (PLSR) was used (www.r-project.org). All

procedures were applied after the missing value estimation followed by normalization of the metabolic matrix.

Network analysis and graph clustering

A matrix of correlation between all trait pairs was generated. Initially the R-package "igraph" was used to visualize the network and then we exported the graph to a file which can be read by DPClus (Altaf-Ul-Amin et al., 2006; Csardi and Nepusz, 2006). Essentially, this algorithm divides the network into modules or groups of vertices that are more connected between themselves than to nodes from others and extract densely connected nodes as a cluster. In this study, we used the overlapping-mode with the DPClus settings since the overlapping-mode is consistent with the overlap of many of the metabolic pathways and protein complexes. The algorithm of DPClus receives three inputs: the network, a value of minimum density we allow for the generated clusters (d_{in}) and a minimum value for cluster property that determines the nature of periphery tracking (cp_{in}). The values for density and cluster properties should be within the following range $0< d_{in} \leq 1$, and $0< cp_{in} \leq 1$ (Altaf-Ul-Amin et al., 2006). We set the parameter settings of cluster property cp; density values were set to 0.5 as it gives the best performance in a graph clustering. All network statistics, such as graph density, were calculated in R with the 'igraph' package).

Results

1- Metabolite Distribution and Detection

We utilized an in-house gas chromatography–time of flight–mass spectrometry (GC-TOF-MS) metabolomics platform to measure metabolite accumulation in the seeds of a *Solanum lycopersicum* (Moneymaker 'MM') *x Solanum pimpinellifolium* ('Pimp') RIL population (Voorrips et al., 2000).

This GC-TOF-MS platform detects predominantly primary metabolites, and metabolites are identified based on comparison with reference spectra (NIST and http://csbdb.mpimp-golm.mpg.de/csbdb/gmd/gmd.html). In total, 160 metabolites were detected in this study and the chemical nature was identified for 66 of these. The known metabolites included central metabolism derived compounds, such as glucose-6-phosphate, members of the tricarboxylic acid (TCA) cycle, such as succinate, citrate and malate, members of the membrane/phospholipid biosynthesis, such as glycerol-3-phosphate, ethanolamine, amino acids and precursors thereof, carbohydrates, and some other common metabolic end products (Supplemental Table S4.1). This list was compiled to encompass the different classes of intermediates in primary metabolism. These metabolites are ubiquitously present in living organisms and are at the core of the biochemical reaction networks with the largest fluxes and largest number of regulatory circuits.

The parents of the RIL population significantly differed in their metabolite accumulation (see Supplemental Figure S4.1). Metabolites differing significantly included branched chain and aromatic amino acids, fatty acids, glutamate and ethanolamine, GABA, myo-inositol, phosphoric acid, carbohydrates and organic acids. Pimp showed higher levels of tricarboxylic acid cycle (TCA) and amino acid–associated metabolites, while MM was higher in galactonate, urea, monomethyl phosphate and GABA. The majority of the metabolites (160) were detected in both parents and in more than 90% of the RILs. The presence or absence of metabolites is likely caused by a mixture of polymorphisms with qualitative and quantitative effects (see Supplemental Figure S4.1). Transgressive segregation for metabolite presence was manifested in a significant fraction of the metabolites found in the RIL population. Analysis of the RILs for 160

metabolites identified positive and negative transgressive segregation for metabolite accumulation (Supplemental Figure S4.2). Thus, *S. lycopersicum* and *S. pimpinellifolium* possess significant genetic variation for metabolite accumulation. This population of 100 RILs was divided in two sub-populations optimized for the distribution of parental alleles using the R-procedure DesignGG (Liet al. 2009); hence 50 RIL lines were used for dry seeds and 50 lines for 6h imbibed seeds.

The data set obtained by GC-TOF-MS was examined by principal component analysis (Figure 4.1A and 4.1B, Table 4.1 and Supplemental Table 4.2). Samples derived from dry and 6h imbibed tomato seeds were clearly separated on the basis of metabolic differences (Figure 4.1B). The analysis further highlighted a clear metabolic shift, in both dry and 6h imbibed tomato seeds. These metabolic events are best exemplified by the metabolites with the highest principal component analysis scores and ANOVA P values in each stage (i.e. those metabolites with the main impact on the variance of the data set; Supplemental Table 4.2). To confirm the most important principal components (PC) of these samples we prepared score plots for the dataset. The loading plots highlighted and visualized metabolites with a significant role in seed developmental stage separation. For example, threonate, pentonic acid and galactinol differed strongly between the dry and imbibed seeds (primarily the first principal component). The complete list of metabolites that contributed to the separation of the dry from the imbibed seeds (discriminative metabolites) is shown in Supplemental Table 4.2. In the PCA score plots of the 66 identified metabolites, the first principal component (PC1) clearly discriminated between dry and imbibed samples (Figures 4.1A and 4.1B). To obtain further insights into the interpretation of our PCA results, we recalculated PCA by removing the three most influencing metabolites

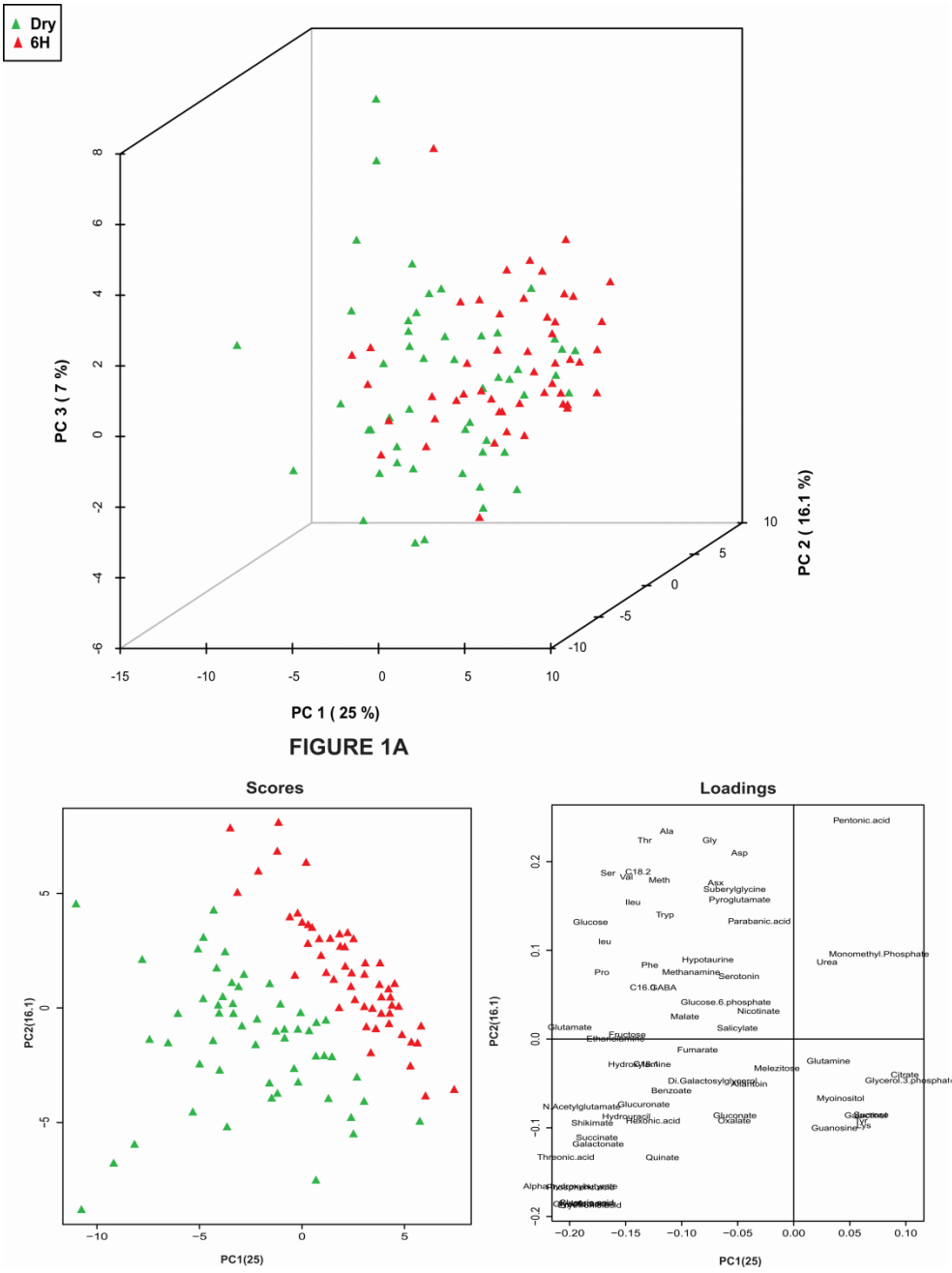


Figure 4.1. Principal component analyses of the 66 commonly detected metabolites for dry and 6h imbibed tomato seeds. Symbols: green

rectangles, dry seeds; red rectangles, imbibed. Each data point corresponds to the analysis of one of 100 genotypes.

(A) 3D-PCA plot for metabolite data

(B) 2D-PCA scores and loadings plots of metabolite data Scores of principal component analysis are presented dry and 6h based on a combination of 2 components (PC1 and 2). Variances 25% for PC1 and 16.1% for PC2 were recorded in each component. Loadings score of metabolites is presented in PC1 and PC2.

(threonate, pentonic acid and galactinol) from the original data matrices. Supplemental Figure 4.3 shows the score and loading plots of dry and 6h imbibed seeds using the remaining 63 seed metabolites. The PCA plots now revealed that glutamate, alanine, and sucrose strongly contributed to PC1 separation.

Quantitative changes in the amounts of major metabolites in the two different stages are presented in Figure 4.2. The progression of seeds from the dry to the imbibed stage was associated with changes in levels of the majority of amino acids and their precursors, alcohols, carbohydrates, organic acids and fatty acids compounds (Figure 4.2). The concentration of organic acids, namely, galactonate, glycolate, glycerate, erythronic acid, quinate and threonate, decreased dramatically upon imbibition. The levels of amino acids and their precursors were did not differ between dry and 6h imbibed seeds. The levels of hydroxybutyrate and the carbohydrates xylofuranose and sucrose also exhibited considerable decrease upon imbibition. The levels of the TCA-cycle intermediate oxalate showed significant decrease while the other TCA-cycle metabolites declined even further on imbibition. The imbibed seed stage was associated with a general increase in concentration of monomethyl phosphate, the organic

Table 4.1. Principle components and ANOVA p-values expressing the statistical significance of the changes in primary metabolites content between dry and 6H tomato seed. Columns represent, respectively, the proportion of variance that could be explained by PC1, PC2, PC3.

Metabolites	PC1	PC2	PC3	Metabolites	Adj. p value
Citrate	0.130827	-0.14777	0.039654	Galactonate	3.29E-32
Monomethyl Phosphate	0.073226	-0.10815	-0.01438	Citrate	6.12E-28
Parabanic acid	-0.02082	-0.12748	0.142907	Threonate	9.54E-28
Suberylglycine	-0.03716	-0.16423	0.026534	Phosphoric acid	4.46E-27
Pentonic acid	0.094462	-0.22537	0.126329	Glycolic acid	6.41E-25
Glycine	-0.05894	-0.23185	0.128291	Xylofuranose	1.19E-19
Pyroglutamate	-0.04282	-0.17416	0.128511	Hydroxybutyrate	8.12E-18
Glycerol-3-phosphate	0.092162	-0.03967	0.206803	N-Acetylglutamate	2.17E-17
Urea	0.028261	-0.08977	0.076872	Erythronic acid	2.73E-16
Nicotinate	-0.03334	-0.04327	-0.08201	Pentonic acid	7.09E-16
Serotonin	-0.03318	-0.06001	-0.06548	Glyceric acid	3.81E-15
Proline	-0.15944	-0.08862	-0.12771	Glutamate	9.75E-14
Valine	-0.12778	-0.18333	-0.17445	Monomethyl Phosphate	7.14E-08
Isoleucine	-0.12351	-0.15227	-0.22146	Hydrouracil	1.18E-07
Leucine	-0.15409	-0.11535	-0.20597	Oxalate	2.66E-07
C18:2	-0.11451	-0.19159	-0.11889	Shikimate	2.14E-06
Tryptophan	-0.1317	-0.19067	-0.09027	Leucine	9.57E-06
Methionine	-0.10234	-0.18519	-0.154	Myo-inositol	2.59E-05
Threonine	-0.1129	-0.23596	-0.03316	Ethanolamine	4.44E-05
Alanine	-0.09558	-0.24485	-1.23E-05	Gluconate	5.21E-05
Serine	-0.1469	-0.20048	-0.02866	Glycerol-3-phosphate	7.90E-05
Aspartate	-0.0373	-0.21633	-0.00462	Succinate	2.41E-04
Asparagine	-0.05577	-0.18233	-0.15608	Hexonic acid	3.78E-04
Hypotaurine	-0.07449	-0.11513	0.026758	Glucose	5.70E-04
Fructose	-0.15339	-0.04639	0.055182	C18:1	9.38E-04
C16:0	-0.14145	-0.09126	-0.06115	Benzoate	9.99E-04
Salicylate	-0.05288	-0.02386	0.025049	Aspartate	1.04E-03
Hydroxylamine	-0.13786	0.006996	-0.02999	Proline	1.24E-03
Sucrose	-0.08454	0.004467	-0.07809	Glycine	1.60E-03
Allantoin	-0.06582	0.014807	-0.12072	Lysine	3.46E-03
Guanosine	-0.04927	0.046388	-0.08514	Glucuronate	4.26E-03
Gluconate	-0.09144	0.064455	-0.17625	Tyrosine	4.48E-03
Hexonic acid	-0.13967	0.054328	0.062403	Isoleucine	6.08E-03
Oxalate	-0.05654	0.098249	-0.18727	Hydroxylamine	7.04E-03
Galactonate	-0.1717	0.114441	-0.15487	Quinate	8.30E-03
Phosphoric acid	-0.18928	0.153397	-0.08304	Pyroglutamate	9.18E-03
Threonate	-0.20214	0.112749	-0.0635	C16:0	1.88E-02

Table 4.1 *Continued*

Metabolites	PC1	PC2	PC3	Metabolites	Adj. p value
Hydroxybutyrate	-0.19861	0.141517	-0.0147	Galactinol	1.94E-02
Xylofuranose	-0.18769	0.164195	-0.04544	Valine	2.19E-02
Glycolic acid	-0.1918	0.166886	-0.06776	Fructose	3.00E-02
Erythronic acid	-0.18213	0.166799	-0.02653	Phenyalanine	1.19E-01
Glyceric acid	-0.18437	0.16335	-0.0183	C18:2	1.63E-01
Fumarate	-0.07746	0.002719	0.26859	Alanine	2.09E-01
Glucose-6-phosphate	-0.07319	-0.0668	0.127507	Di-Galactosylglycerol	0.031465
GABA	-0.11896	-0.0871	0.109405	Guanosine	0.037789
Ethanolamine	-0.16154	-0.03853	0.167042	Serine	0.043951
Glutamine	-0.07775	-0.11774	0.174113	Allantoin	0.048688
Glucose	-0.16719	-0.1324	0.079396		
Myo-inositol	-0.18784	-0.0457	0.041628		
Shikimate	-0.16909	0.095877	0.091835		
Glutamate	-0.1964	-0.04181	0.017139		
N-Acetylglutamate	-0.19283	0.04814	0.126887		
Quinate	-0.11327	0.110813	0.153603		
Succinate	-0.16756	0.089067	0.16658		
Benzoate	-0.10212	0.052636	0.085576		
Glucuronate	-0.1218	0.066618	0.287764		
Malate	-0.09287	-0.04158	0.230577		
Melezitose	-0.00894	0.039725	0.126793		
Phenylalanine	-0.11829	-0.09801	0.068139		
Methanamine	-0.08836	-0.09114	0.093967		
C18:1	-0.11776	0.031624	0.189357		
Lysine	-0.03974	0.087368	-0.03871		
Hydrouracil	-0.15544	0.067386	-0.0361		
Tyrosine	-0.04627	0.070562	-0.04948		
Di-Galactosylglycerol	-0.06457	0.051115	0.215632		
Galactinol	-0.10889	0.014501	0.070099		

acids, parbanic acid and pentonic acid, and the TCA-cycle intermediates citrate and fumarate. In contrast, the levels of gluconate, quinate, shikimate and succinate were significantly reduced. While the levels of most amino acids and their precursors were reduced to different extents, the levels of glycine, aspartate, asparagine and hypotaurine significantly increased. Similarly, the levels of most sugars declined but the levels of the sugar phosphates glucose-6-phosphate and glycerol-3-phosphate were elevated significantly. This general observation suggests that the transition

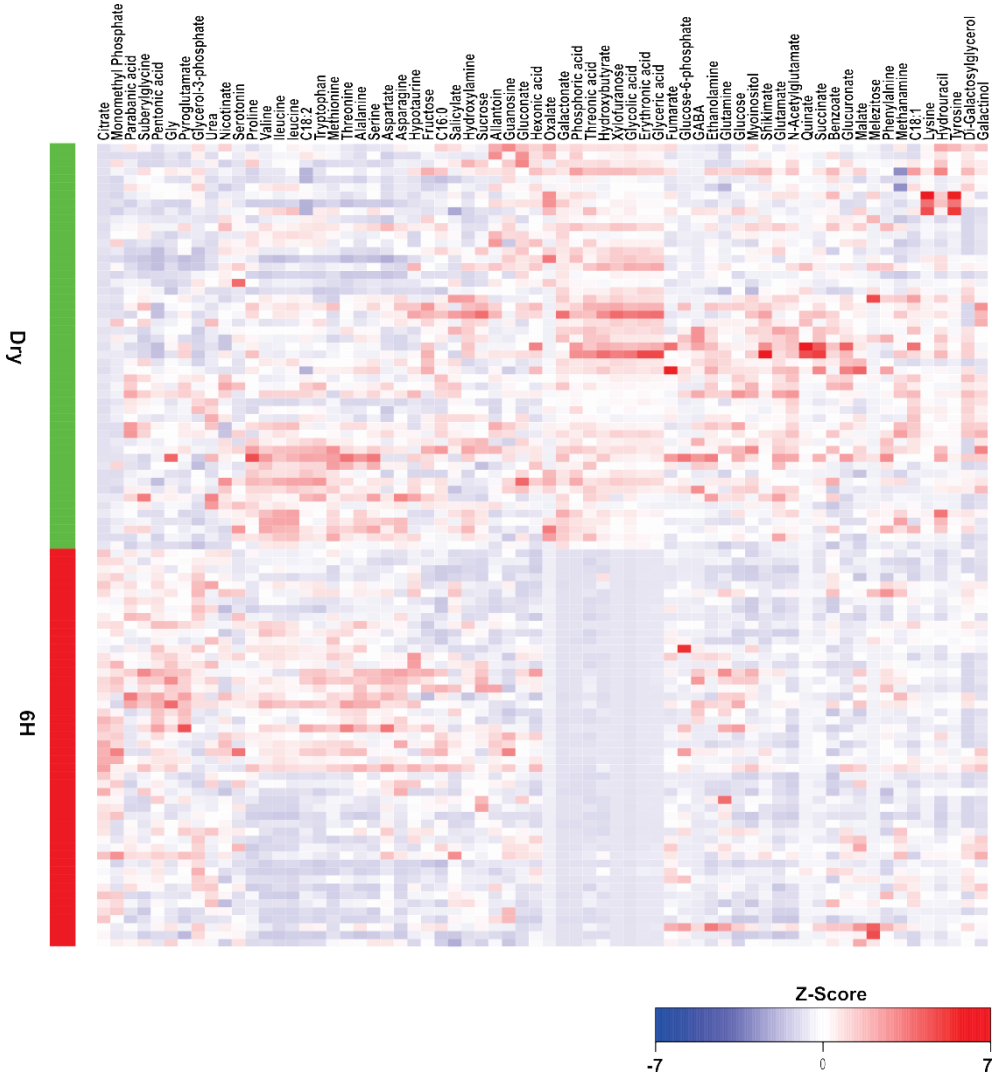


Figure 4.2. Metabolite profiles in dry and 6h imbibed seeds of tomato. The vertical Green and Red bar colors represent variability of metabolite abundance between F_8 recombinant inbred lines (RILs) in dry and 6h imbibed seeds. The relative abundance of each metabolite represents the mean of all genotypes (n = 101). A hierarchical classification of metabolites was done according to a dissimilarity scale using the distance function 1-

correlation. The dissimilarity index is employed for cluster analysis to arrange different metabolites according to their similarity (Legendre and Legendre, 1998). Z-values of measurements are color-coded as indicated in the scale on the bottom of the heat map, from blue to red.

from dry to 6h imbibed is associated with the activation of initially important metabolic processes needed for seed germination. It is also likely that germination is associated with a follow up of additional metabolic processes, which occur later during germination and therefore were not observed by our metabolic profiling.

2- Metabolites of Similar Function are Highly Correlated across the F_8 Population

We created a correlation matrix of all pairwise comparisons among individual metabolites by performing Spearman rank correlation analysis for all pairs of measured traits across the whole population. Spearman's rank correlation coefficients (Rs) and accompanying false discovery rate (FDR)–corrected P values (p_{BH}; Benjamini-Hochberg) are provided in Supplemental Table 4.3. Unsupervised hierarchical clustering revealed several "hot spots" of highly correlated metabolites (Figure 4.3, Supplemental Figure S4.4). It is remarkable that several hot spots corresponded to the biochemical pathways to which the metabolites belong. For example, 11 of the 15 amino acids cluster together. Moreover, when we consider pairwise correlations between all amino acids, 75% had absolute correlation coefficients greater than R_s 0.38 (p_{BH} = 0.0001)(Supplemental Table 4.3). In another cluster, glycine clusters most highly with pyroglutamate (Rs = +0.64; p_{BH} = 1.36E-11), but also with glycerol-3-phosphate and urea. In most organisms the major thiol is the tripeptide glutathione (γ-Glu-Cys-Gly, known as GSH) and its utilization is linked by the γ-glutamyl cycle (Ritz and

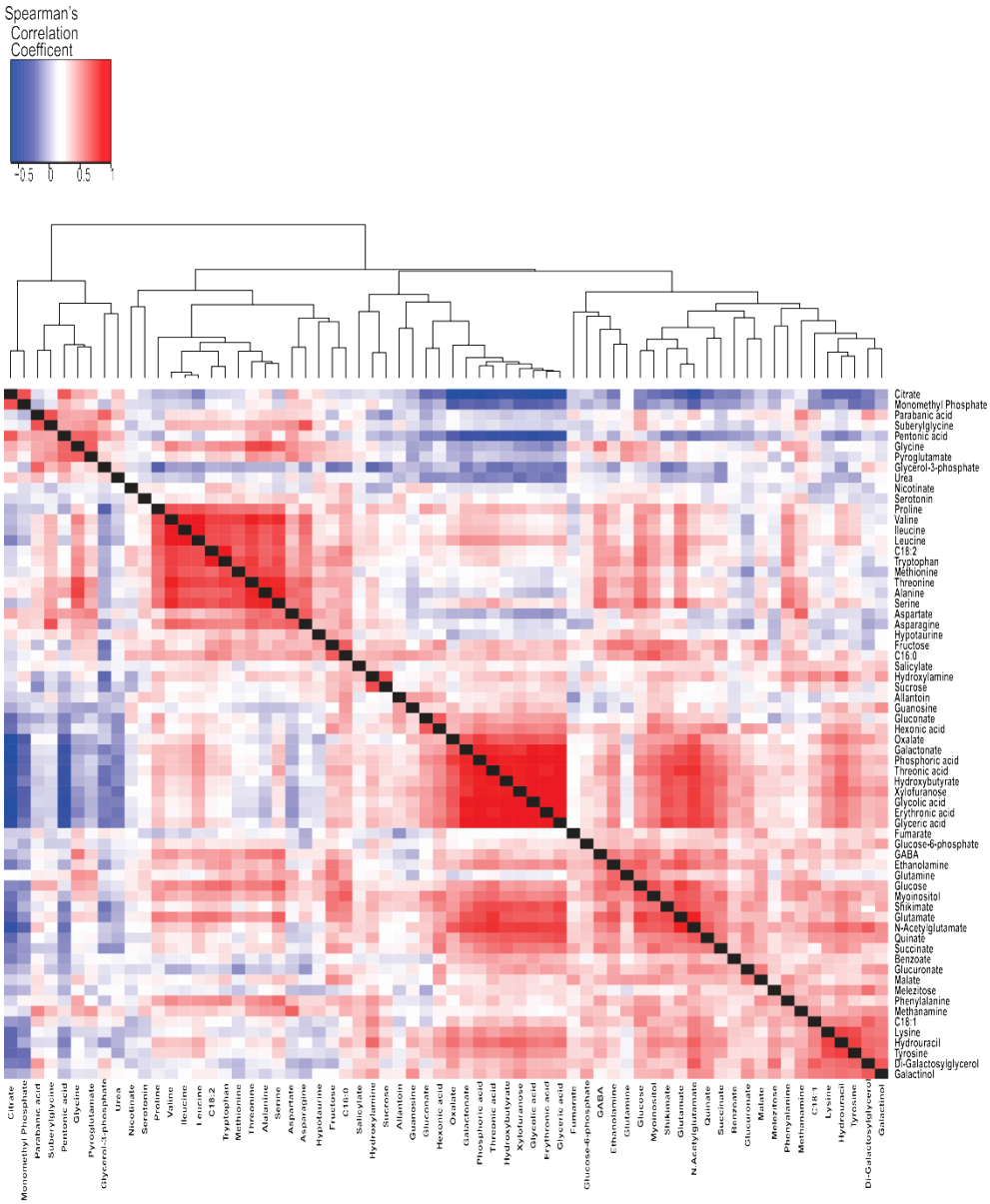


Figure 4.3. Heat map of correlations between metabolites. Each square represents the Spearman correlation coefficient between the metabolic phenotypes of the column with that of the row. Metabolic phenotype order is determined as in hierarchical clustering using the distance function 1-correlation. The dissimilarity index is employed for cluster analysis to

arrange different seed phenotypes according to their similarity (Legendre and Legendre, 1998). Self-self correlations are identified in black. Individual correlation coefficients can be found in Supplemental table S4.3. Supplemental Figure S4.3 displays the correlation heat map for all 160 metabolites found in our analysis.

Beckwith, 2001). Glutathione synthetase catalyzes the reaction glycine and the cysteine carboxyl of the L-γ-glutamylcysteine dipeptide to form GSH (Ullmann et al., 1996; Wang and Oliver, 1996; Jez and Cahoon, 2004). During glutathione degradation the breakdown products include glutamate, glycine, and cysteine which can be reabsorbed into the cell and used for additional GSH synthesis. Glycine is classified as a glucogenic amino acid since it can be converted to serine by serine hydroxymethyltransferase, and serine can be converted back to the glycolytic intermediate, 3-phosphoglycerate or to pyruvate by serine/therionine dehydratase (Datta et al., 1987; Ogawa et al., 1989 ; Wang and Ballatori, 1998). Nevertheless, the main glycine catabolic pathway leads to the production of CO_2 and ammonia. Nitrogen waste originating from protein and amino acid catabolism can be substrate for the urea cycle.

The fact that metabolites of a common functional group are highly correlated suggests that there are potential regulators of these biochemical pathways segregating in this F_8 RIL population. Glucose correlates most highly with myo-inositol (Rs = +0.68; p_{BH} = 9.06E-05), and GABA with glutamate (Rs = +0.58; p_{BH} = 1.25E-09). Myo-inositol (MI) is a ubiquitous compound synthesized from D-glucose in three steps: firstly, glucose is phosphorylated by hexokinase; secondly, glucose-6-P is converted to 1L-myo-inositol-1-P by 1L-myo-inositol-1-phosphate synthase, and, finally, 1L-myo-inositol-1-P is dephosphorylated by a phosphatase to produce free MI.

In plants, myo-inositols play various roles as they appear to be involved in phosphate storage in the form of phytic acid (a hexaphosphorylated form of MI), cell wall biogenesis, control of auxin physiology, membrane biogenesis, signal transduction and stress tolerance (Meng et al., 2009). Also, in the same cluster ethanolamine, glutamine shikimate and its precursor quinate, as well as TCA intermediates, malate and succinate, group together. Decarboxylation of glutamate (Glu) produces GABA and CO_2 in the cytosol and presumably it is transported to the mitochondria where it is converted to succinic semi-aldehyde. Subsequently, succinic semi-aldehyde is converted either to succinate or 4-hydroxybutyrate.

3- Metabolomic QTL (mQTLs) Location

To test if there is genetic variation affecting tomato seed metabolic composition we used the generalized genetical genomics model (GGG) for the analysis of quantitative variation for seed metabolites for genetic factors or genetic times controlled environmental perturbations in combination (Li et al., 2008). The generalized strategy enabled us to map quantitative trait loci (mQTLs) due to genetic effects as well as to detect how QTL effects differ across dry and imbibed environments and how the genotype influences the response to environmental changes in genetic-by-environment QTLs (G x E QTLs). We did QTL mapping using a linear model: Y = G + GxE + E (where Y is the matrix of metabolite measures) and separate –log10 probability values (called LOD scores hereafter) are generated for the environmental (E), genetic (G) and genetic x environmental (G x E) linkage. Significant thresholds were defined with permutation analysis (n=1000, $p<0.01$) by randomizing the genotypes over each metabolite and they were set to LOD >3 accordingly. We found that individual metabolites mapped to specific regions of the genome (Supplem-

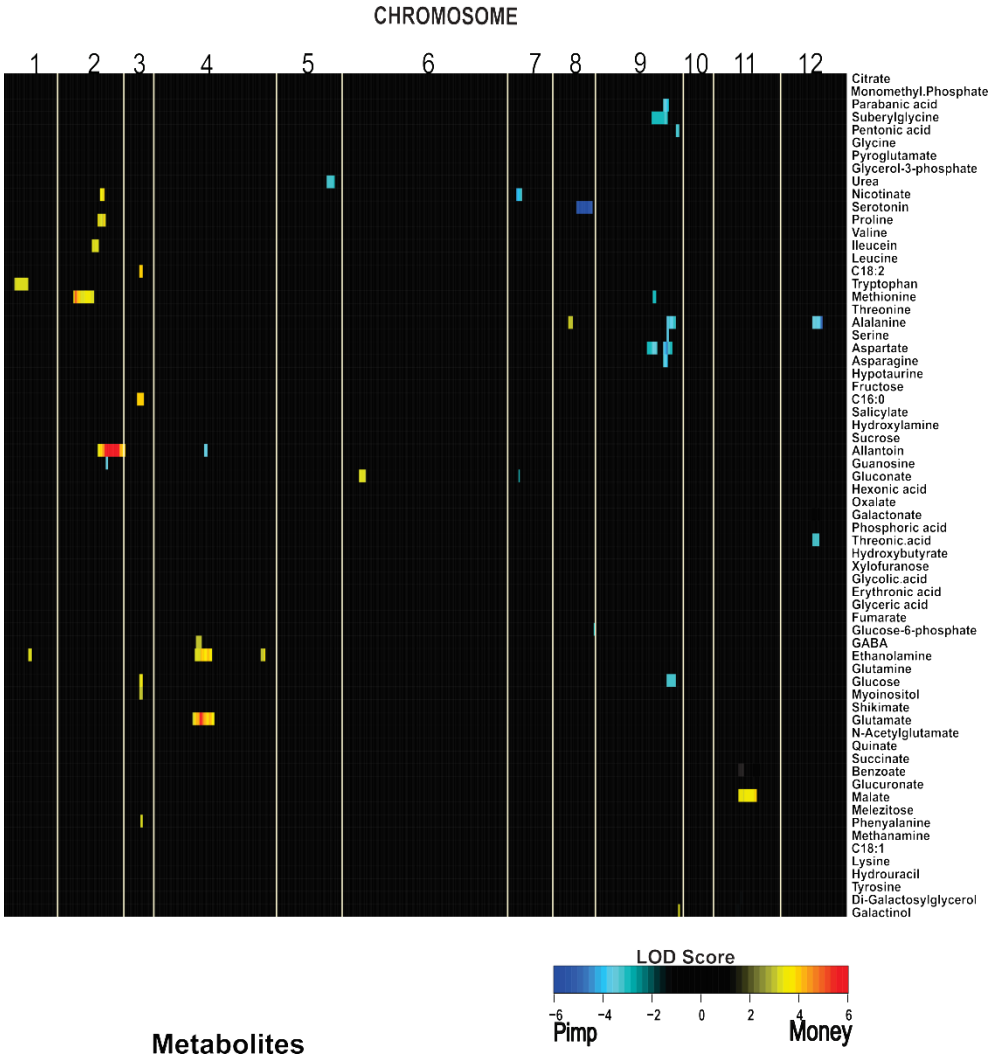


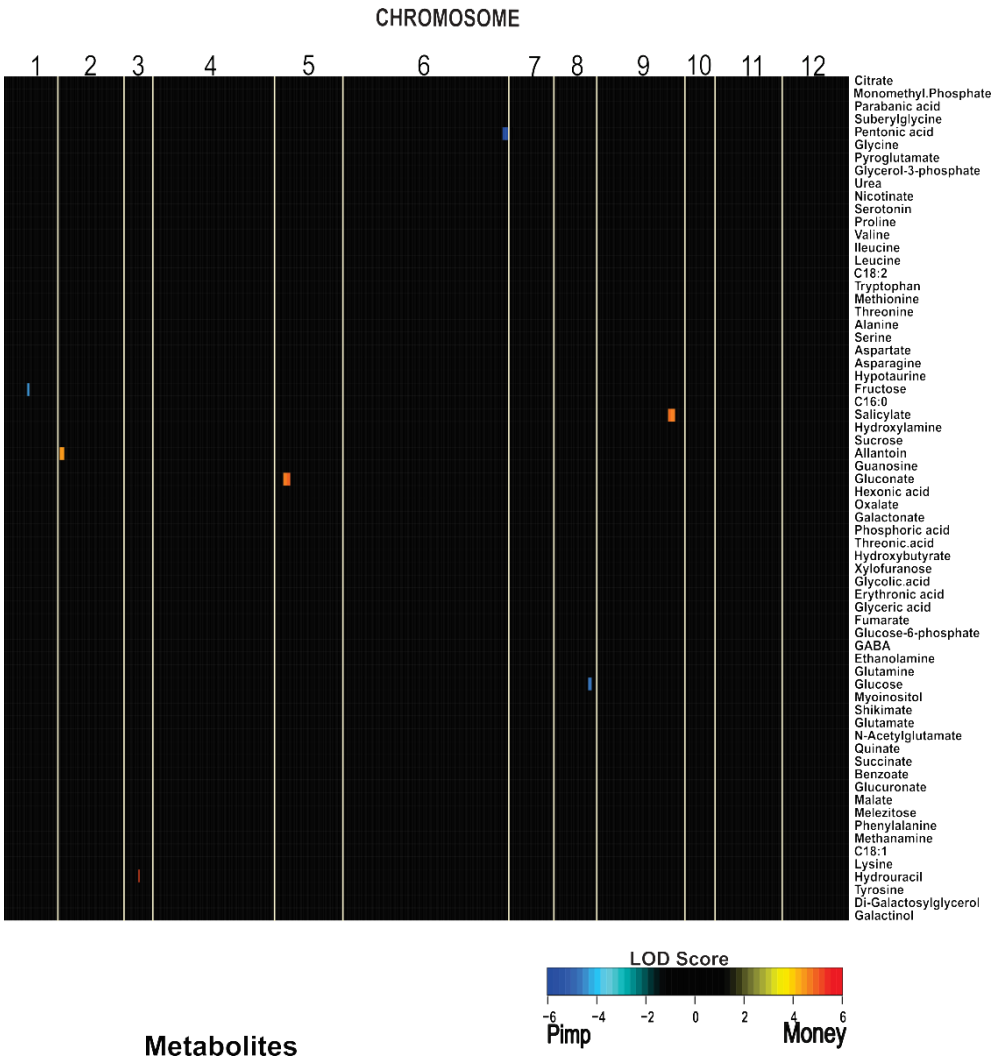
Figure 4.4. (A)- Genomic locations of genetic mQTLs identified for metabolite accumulation. Tomato chromosomes are identified by roman numerals (1–12), with centimorgans ascending from left to right; chromosomes are separated by white lines. Colored cells indicate QTL significant. Significant thresholds were defined with permutation analysis (n=1000, $p<0.01$) by randomizing the genotypes over each metabolite and was set to LOD >3 accordingly. The LOD color scale is indicated, showing

blue and light blue when the *Solanum pimpinellifolium* ('Pimp') allele, and yellow and red when the *Solanum lycopersicum* (Money 'MM') allele, at that marker results in an elevated level of metabolic phenotype. QTL positions, LOD scores and effects are provided in Supplemental Table S4.4. Supplemental Figure S4.4A and S4.4B shows genomic locations of genetic mQTLs identified for all 160 metabolites.

ental Table 4.4, Figure 4.4A and 4.4B, Supplemental Figure S4.5A and 4.5B). By R/qtl analysis, using the linear model, we detected mQTLs for more than 50% of the metabolites, indicating the genomic regions that influence (either directly or indirectly) the abundance of these metabolites. Figures 4.4A and 4.4B display heat maps of genetic and genetic-by-environment interactions (G x E QTL) mQTLs, respectively, with metabolites organized by hierarchical clustering (Supplemental Table 4.4, Figure 4.4A and 4.4B, Supplemental Figure S4.5A and 4.5B). In our experimental set-up the environmental variation is defined as variation observed between the 2 developmental stages (dry and 6h). Although a concentrations of metabolites, we also found metabolites with variation in the genetic xenvironmental (G x E QTL) component (Supplemental Table 4.4, Figure 4.4A and 4.4B, Supplemental Figure S4.5A and 4.5B).

Co-location of the QTLs was expected, as there was strong correlation among metabolites, which is an indication of possibly shared mQTLs. Four of the total eleven amino acids that clustered based on correlation map to a similar position on chromosome 9 (Supplemental Table 4.4, Figure 4.4A and 4.4B, Supplemental Figure S4.5A and 4.5B). Glutamate and GABA have a QTL profile with a shared mQTL on chromosome 4 (Figure 4.3). Co-location of mQTLs is assumed to be associated with co-regulation of metabolites and may indicate that a

specific biological function controls different components or that a specific large fraction of the observed variation is due to genetic effects (G) among step in a biochemical pathway is affected. QTL mapping, using a linear model: Y = G + G x E + E on the data of individual environments, led to



(B)- Genomic locations of G x E mQTLs identified for metabolite accumulation.

the identification of only a few mQTLs for genetic-by-environment interaction on all 12 chromosomes of tomato (Figure 4.5, Supplemental Figure S-4.5). One possible reason could be the size of the mapping population in our analysis as mapping is strongly dependent upon the population size and recombination frequency (Mackay et al., 2009). It corroborates with the findings of Keurentjes *et al.,* (2007) who showed linear relationship between the number of individuals used for mapping and the detectable genetic effect.

Integration of metabolic and seed phenotypic traits

Using an analogous approach to that taken previously for metabolites (Meyer et al., 2007; Sulpice et al., 2010), we were also interested in unravelling possible links between previously studied seed quality phenotypes and a specific combination of metabolites. These seed quality parameters are: maximum germination (G_{max}, %), the onset of germination [t_{10}^{-1}; reciprocal of time to 10% of germination of viable seeds (h^{-1})], the rate of germination [t_{50}^{-1}; reciprocal of time to 50% of the germination of viable seeds (h^{-1})], MGR = mean germination rate, which is the reciprocal of the mean germination time (MGT^{-1}), uniformity (U_{7525}^{-1}, reciprocal of time interval between 75% and 25% viable seeds to germinate; h^{-1}) and area under the germination curve [AUC; the integration of the fitted curve between t = 0 and 200 h] under different germination conditions (Kazmi et al., 2012). The correlation matrix was constructed using Spearman's rank correlation and it contained 160 metabolites and 42 seed quality phenotypes, allowing 17,205 pair-pair correlations between seed quality phenotypes and metabolites (Supplemental Table 4.5). We only found metabolites that correlated weakly with germination traits across non-stress (control), as well as

different stress, conditions (see Supplemental Table 4.4). We therefore used canonical correlation analysis (CCA), which is a multivariate technique, and previously it has been shown that a canonical combination of metabolites can be used to predict certain phenotypic traits (Meyer et al., 2007; Lisec et al., 2008; Sulpice et al., 2010).

We extracted the metabolites most relevant for germination traits by their correlation to the canonical variate. The most striking result to emerge from the analysis is that there is no single metabolite associated with germination parameters. The first 20 metabolites with significant correlations under oxidative stress condition (G_{max}) are listed in Table 4.2 as an example, whereas details concerning correlation between metabolites and different germination traits are provided in Supplemental Table 4.5. The highest absolute correlation found was for an unknown metabolite (RI_2442), which yielded a value of 0.406. Although the correlation is stati istically highly significant (P value of 3.35E-05), it can only explain 16.48% of the variance. Other significantly correlated compounds with G_{max} are allantoin, hydroxybutyrate, C16:1, fructose, gluconate, glucuronate, guanosine, hexonic acid, hypotaurine, shikimate, xylofuranose and a number of unknown metabolites (Supplemental Table 4.5). Their individual contribution to the explained variance ranges from 5%-15%. In contrast to the aforementioned pairwise correlation analysis, CCA yielded a much stronger correlation of 0.60. This value corresponds to 36% of variance explained by a linear combination of metabolites, almost 1-5 times more than explained by any individual metabolite. Comparing the results obtained from a combination of different metabolites and germination traits which showed significant associations, it can be seen that strongly represented metabolites are compounds of central metabolism, such as glucose and fructose, members of the tricarboxylic acid (TCA) cycle, such

Table 4.2. List of top 20 signature metabolites ranked according to the strength of the canonical correlation (G_{max} Oxidative Stress; 300mM H_2O_2).

Metabolites	COR	PV
RI_1379	0.550394	4.33E-09
Fructose	0.549864	4.51E-09
RI_1957	0.548966	4.84E-09
RI_2866	0.537668	1.14E-08
Hexonic acid	0.53543	1.34E-08
RI_2442	0.531606	1.78E-08
Hypotaurine	0.479808	5.76E-07
Hydroxybutyrate	0.468234	1.16E-06
Shikimate	0.467711	1.20E-06
RI_2882	0.465292	1.38E-06
RI_2470	0.434414	7.85E-06
RI_1867	0.434402	7.85E-06
Glucuronate	0.405841	3.38E-05
RI_2939	0.403708	3.75E-05
RI_1703	0.395356	5.60E-05
C16:0	0.383011	9.92E-05
Allantoin	0.359124	0.000282
RI_2087	0.353745	0.000353
Xylofuranose	0.347831	0.000449
Guanosine	0.286496	0.004237

as succinate, citrate and fumarate, amino acid and precursors, members of the membrane/phospholipid biosynthetic pathways, such as glycerol-3-phosphate, ethanolamine and myo-insitol, or sucrose (Supplemental Table 4.5). Interestingly, with regard to different germination traits ($G_{max,}$ t_{10}^{-1}, t_{50}^{-1}, MGR, and AUC), in particular under stress conditions, the predominantly represented metabolites are well known for abiotic stress responses such as myo-insitol, proline, fumarate and succinate. High temperature stress was associated with isoleucine, leucine and valine; again their response to abiotic and biotic stresses and heat is well known, whereas strong representation of carbohydrates (xylofurnose, fructose), organic acids (gluconate, glucornate, hexonate, shikimate etc.) and alcohols was evident in the case of G_{max} and some additional metabolites in the case of AUC

under oxidative stress condition. The overall association of metabolites with phenotypic traits across different conditions provides an indication of possible cross-talk between different abiotic stresses (e.g. salt and osmotic).

6- Co-localization of phQTLs with mQTLs

One of our aims of the present study was to link mQTLs with previously found phenotypic QTLs (phQTLs) (Kazmi et al., 2012). Inspection of the overlap showed that several QTLs controlling primary metabolites were co-located with different seed quality phenotypes (Supplemental Figure 4.6). A comprehensive overview of all overlapping mQTLs with phQTLs observed in the RIL population for known metabolites and the chromosomal localization is shown in Supplemental Figure 4.6 and Supplemental Table 4.6 with the number of overlapping mQTLs per phQTL ranging from 3 to 9. We therefore used a permutation test to identify statistically significant overlaps. This indicates that there is strong genetic regulation of the metabolic and phenotypic traits investigated in this study and also points at possible cross-talk in different seed germination conditions. Some metabolites (allantoin, pentonic acid, monomethyl phosphate, melezitose etc.) display up to two QTLs co-localizing with any of the phQTL (Supplemental Figure 4.6).

7- Construction of a metabolomic correlation network

An overview of the metabolic correlations network was obtained by extracting all significant trait-trait correlations ($R_s \geq 0.5$) and visualizing them using DPClus (Altaf-Ul-Amin et al., 2006) that identifies clusters in the metabolic correlation network. The selection of an appropriate threshold for

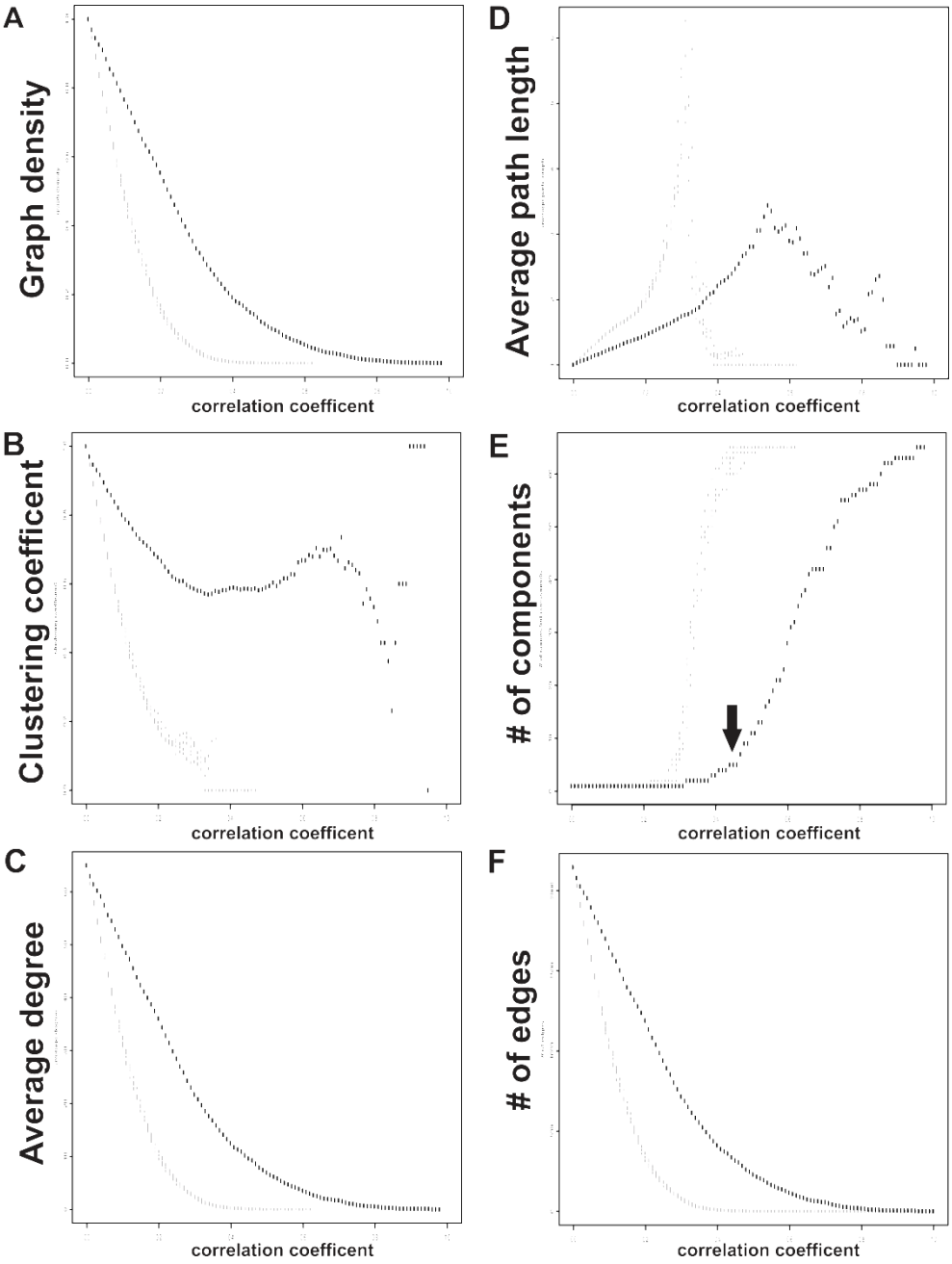


Figure 4.5. Correlation network properties of the 66 metabolites in the tomato seeds across a range of correlation coefficients. Networks were

constructed for a range of correlation thresholds from 0 to 1.0 by 0.01 increments, and each resulting network was calculated for: (A) the graph density - the ratio of the number of edges and the number of possible edges, (B) the clustering coefficient, (C) the average degree of all nodes, (D) the average path length, (E) the number of connected components, and (F) the number of metabolite-metabolite correlations (edges) in the network. Within each plot, black solid circles represent the observed data points; black dots represent 100 randomized data.

the construction of such a correlation network is accomplished by following methods described by Fukushima *et al.*, (2011). We set the threshold at $R_s \geq 0.5$ by assessing threshold-dependent changes in the network topology, i.e., the graph density, clustering coefficient, average degree, average path length, number of connected components, and number of edges (Figure 4.5, Supplemental Table 4.6). Although a threshold does not guarantee explicit biological significance, we believe that examining various statistics among the constructed networks is useful for selection. A full list of KEGG pathways used in this study is shown in Supplemental Table S4.6. We were able to identify 6 clusters in the metabolomic correlation network of tomato seed metabolites ranging in size from 4 to 16 metabolites (Figure 4.6). To assess the significance of the clusters we used the over-represented KEGG pathways approach (so-called KEGG enrichment analysis) analogous to enrichment analysis are presented in Figure 4.6 and enriched KEGG pathways included glyoxylate and dicarboxylate metabolism ($p = 0.00095$), aminoacyl-tRNA biosynthesis ($p = 0.00003$), glycine, serine and threonine metabolism ($p = 0.01013$) and galactose metabolism' ($p = 0.03160$) (Figure 4.6). The statistical relevance of this approach was evaluated with the S-value (Alexa et al., 2006; Fukushima et al., 2011) that

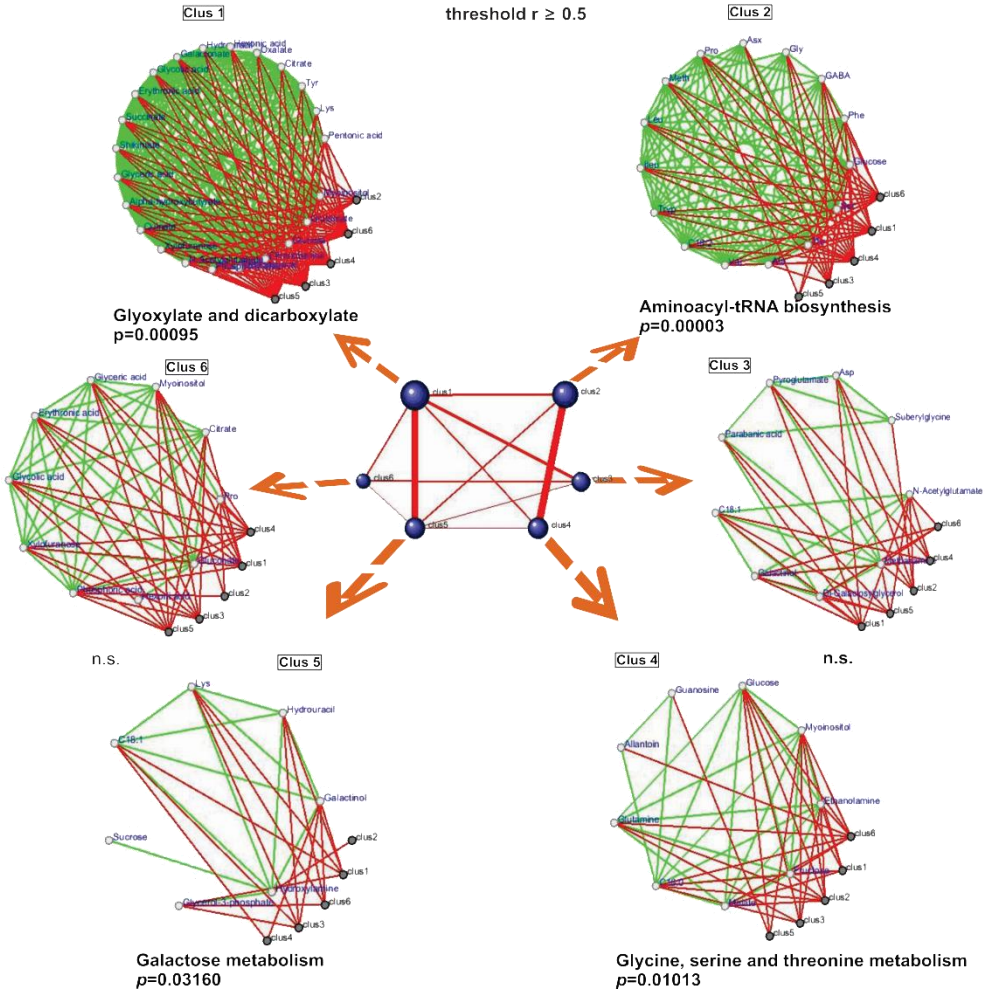


Figure 4.6. Graph clustering of correlated metabolomic modules in tomato seeds (threshold r ≥ 0.5). Using the DPClus algorithm we extracted 6 clusters in tomato seeds. The significant metabolic pathways were assigned by KEGG enrichment analysis (see Methods). The central graph consisting of 6 blue clusters and 10 red edges was extracted by DPClus. Each blue cluster contains densely connected metabolites (see Clus1 to 6). Small white nodes in the clusters indicate metabolites. The internal nodes of the clusters are connected by green edges; neighboring clusters are connected by red edges.

can be used for assessing the significance of the clustering results based on KEGG pathways (Fukushima et al., 2011). The strong correlation between amino acids confirmed the observations as described above. Metabolites belonging to amino acid biosynthesis tended to cluster together. The largest cluster for the seeds was 'glyoxylate and dicarboxylate metabolism and aminoacyl-tRNA biosynthesis'.

Discussion

Approaches employing transcriptomics, proteomics, and metabolomics have yielded vast data sets, allowing the correlation of physiological states with patterns of gene expression, protein levels, and metabolite abundance. A major challenge in the analysis and interpretation of this data is delivering models of causation from these data sets. Progress made in analytical and statistical techniques now enables the construction of regulatory networks that integrate the different levels of biological information, including transcriptional and (post) translational regulation as well as metabolic signalling pathways.

Coordinated changes of metabolites that are required for metabolism in dry and 6h tomato seeds

In the dry stage we observed higher levels of many metabolites, including organic acids, carbohydrates, and levels of alcohols such as hydroxybutyrate, as compared to 6h imbibed seeds. Most prominent were oxalic acid, glycolate, threonate, glycerate and erthyryonic acid. Synthesis of oxalic acid is accomplished via several pathways. Glucose, acetate and some acids of the TCA cycle were determined to be involved in oxalate synthesis (Chang and Beevers, 1968). Moreover, glycolic and isocitric acids (Millerd et al., 1963, 1963) and oxaloacetic acid (Chang and Beevers, 1968)

are known to donate carbon to oxalic acid in plants. Oxalate may be produced by plants through the action of lactate dehydrogenase on glyoxylate (Davies and Asker, 1983; Libert and Franceschi, 1987). Two glyoxylate molecules dismutate to yield glycolate and oxalate. The distribution and activity of dismutase enzyme in oxalate-accumulating species have not been determined, and it is difficult to assess the contribution of this pathway to total oxalate contents in the plant. There are various sources of glyoxylate in higher plants (Libert and Franceschi, 1987). Glycolate is formed in the chloroplast during photorespiration and is subsequently converted to glyoxylate in the peroxisomes (Tolbert, 1981). Another pathway of oxalic acid formation, not involving glyoxylate, is the metabolic transformation of L-ascorbic acid (Saito, 1996; Ilarslan et al., 1997). Cleavage of the ascorbic acid molecule can produce oxalic acid and L-threonate. This conversion does not involve glycolate, glyoxylate, or glycolate oxidase. Ascorbic acid synthesis takes place in plastids. Legume tissue active in nitrogen assimilation can condense CO_2 with phosphoenol pyruvate forming oxaloacetate, which stimulates both the TCA cycle and ammonia assimilation (Ilarslan et al., 1997). Thus, the observed transient accumulation of oxalate in seed tissues could be associated with ureide degradation and subsequent amino acid synthesis, which is required for seed storage protein synthesis in developing seeds. It has been proposed that oxalate degradation is developmentally regulated and has important roles in plant development, including cell wall biochemistry, tissue remodeling (Lane, 1994) and salt stress and homeostasis (Hurkman and Tanaka, 1996). Also, the Ca^{2+} and H_2O_2 released as a result of oxalate degradation could possibly play a role in plant signal transduction (Dixon et al., 1994). One additional possibility is that during seed development there is a change toward a more acid tissue pH, which would allow for the

dissolution of the crystals and release of the Ca^{2+} and oxalate ions. It has been hypothesized that degradation of organic acids in seed development may provide the energy needed for metabolic activity in this period (Ilarslan et al., 1997).

The major metabolic changes observed after 6h of imbibition were a significant reduction in the levels of the majority of metabolites, which had accumulated in the dry seeds. Yet, our present finding suggests that metabolism during the 6h seed stage has an additional function: to render certain metabolites rapidly available to support metabolic recovery during imbibition. This implies that primary metabolites might be rapidly consumed to support the metabolic switch toward enhancing biosynthetic processes needed for early germination. In higher plants, the primary seed storage reserve often are triacylglycerol (Beaudoin and Napier, 2004). During seed germination, triacylglycerol is broken down into fatty acids and glycerol. Fatty acids are converted to sucrose via the glyoxylate cycle to support seedling growth. Glycerol is phosphorylated to glycerol-3-phosphate by glycerol kinase, and then converted to dihydroxyacetone phosphate by glycerol-3-phosphate dehydrogenase (Shen et al., 2003). In the second pathway, glycerol is converted to dihydroxyacetone by NAD-dependent glycerol dehydrogenase, and then phosphorylated to dihydroxyacetone phosphate by dihydroxyacetone kinase. Glycerol kinase and glycerol-3-phosphate dehydrogenase activities have been detected in germinating seeds. The highest expression of the genes encoding these enzymes was detected during seed germination (Eastmond, 2004). The Asp-family pathway results in the synthesis of the essential amino acids lysine, threonine, methionine and isoleucine through several different branches (Galili, 2011). In addition, threonine is also metabolized to glycine, which is involved in plant photorespiration whereas isoleucine is a well-documented

donor metabolite that feeds the TCA cycle in plants. This general expression behavior of amino acid metabolism operates as part of a comprehensive program to suppress biosynthetic pathways in order to preserve the existing energy and stimulate catabolic pathways to generate additional energy and exposes a significant regulatory metabolic link between the Asp-family pathway and the TCA cycle, whose biological function may have a major impact on the physiological response of plants to various abiotic stresses that cause energy deprivation (Baena-González and Sheen, 2008). In our study we show that early germination (imbibition) events are characterized by the efficient reactivation of metabolic pathways via the availability of key precursors as well as coordination of energy metabolism. Several conserved features are apparent in both seed stages analyzed, thus confirming a high biological relevance of these changes in the process of seed and seedling development.

Correlations and mapping of metabolites (mQTLs)

The purpose of the current study was to explore the possibility that the levels of metabolites in tomato seeds are sufficiently heritable to provide significant linkage signals, allowing the identification of mQTLs. Given that many pathways converge upon common metabolites and that these pathways have multiple controllers, any one genetic locus may not alter metabolite levels significantly, and therefore may not be identified as an mQTL. Nonetheless, in our F_8 population, we found significant linkage signals, including some that are quite strong (e.g. allantoin: LOD 7.37, $p<0.01$; chromosome 2). This study also provides evidence of co-regulation of biologically-related pathways. An example of this is the correlation between levels of amino acids and other metabolites and precursors

(Figure 4.3). These findings are consistent with our understanding of metabolic physiology. Seed germination is associated with degradation and mobilization of the reserves that are accumulated during maturation (Bewley, 1997; Borisjuk et al., 2004; Penfield et al., 2005). The efficiency of reserve mobilization during germination, and, consequently, of seedling establishment itself, apparently depends on the extent of reserve accumulation during seed maturation. Early germination events are characterized by the efficient reactivation of metabolic pathways via the availability of key precursors as well as a coordination of energy metabolism (Figure 4.2). Our results reveal that metabolites can be mapped to distinct genetic regions, much like mRNA transcripts (eQTL). The amino acid metabolites provide the most striking evidence of functional clustering. We see in both the correlation matrix (Figure 4.3) and the genetic linkage data (Figure 4.4) that most of the amino acids group together. However, a number of the amino acids, namely methionine, uniquely mapped to chromosome 2, whereas alanine and serine mapped to 8 and proline, aspartate and tryptophan, to chromosome 1. There is an overlap between GABA and glutamate on chromosome 4. This unique GABA/glutamate mQTL may explain why GABA clusters with glutamate rather than the amino acids in the correlation matrix. Glutamate is a substrate and product in amino acid catabolic and biosynthetic pathways. It can act either as an ammonium donor or acceptor in transamination reactions (via α-ketoglutarate) and the glutamate dehydrogenase reaction, and can also be rapidly synthesized from glutamine by glutaminase, thus providing precursor metabolites for the generation of other organic acids and amino acids. This is an indication of co-regulation of the glycolytic events and other metabolic processes occurring during imbibition, e.g. amino acid metabolism. Glutamine can also act as a signalling molecule to

alter expression of the urea cycle and gluconeogenic enzymes. Amino acids closely related by a biochemical pathway exhibited even stronger correlations indicating that ratios between amino acid levels within a seed must be maintained (Toubiana et al., 2012). This reflects a highly-regulated amino acid metabolism that includes both protein and non-protein amino acids (i.e. GABA) and both aromatic and aliphatic, likely to occur at the post-transcriptional level in the regulation of N allocation. However, integration of induced changes at the transcriptional level accounting for the intragenotypic correlation of amino acid metabolism cannot be ruled out. The clustering of these metabolites based on their relative concentration in the F_8 population suggests that static metabolic profiling can be used as a marker for changes in flux through certain metabolic pathways. Our data predict that these metabolites are driven by different genetic regulators, leading to a unique mapping signature, even within a group of highly correlated metabolites. In addition, a number of mQTLs mapped to multiple positions, which indicates complex regulation.

Integration of metabolic and seed quality phenotypic traits

Our study provides a large matrix of 160 metabolic traits, which can be analyzed to identify metabolic traits that are linked with seed quality phenotypes. Multivariate statistics have already been applied to predict germination parameters from a large set of metabolites to show that integrative biomarkers have a highly significant positive or negative correlation with germination parameters and to capture much of the information present in the metabolite profile (Meyer et al., 2007; Lisec et al., 2008; Buscher et al., 2009; Sulpice et al., 2010). The present study corroborates previous studies in which the levels of a large number of metabolites, rather than a few individual metabolites, show a close

correlation with germination parameters (Meyer et al., 2007; Lisec et al., 2008; Sulpice et al., 2010). It indicates that variation in the germination parameters coincides with characteristic combinatorial changes of metabolite levels, whereas individual metabolites may fluctuate largely independently of alterations in germination. CCA provided highly-ranked clusters in which metabolites of central metabolic pathways are strongly represented. Carbohydrate of high relevance are the glucose, sucrose and fructose. Carbohydrate play an important role in overriding developmental regulation of seeds at a given point in time in a given cell or tissue. Previous studies have provided correlative evidence that certain carbohydrate levels and/or the resulting changes in osmotic values are necessary within defined tissues or cells to maintain a distinct stage of differentiation or to proceed with the developmental program. A high hexose (probably mainly glucose) level seems to maintain the capacity of cells to divide, whereas - later in seed development - a certain sucrose level is necessary to induce storage-associated cell differentiation (Wobus and Weber, 1999). An apoplastic invertase preferentially expressed in the inner cell layers of the seed coat is a determinant of hexose levels in legume seeds (Wobus and Weber, 1999). This enzyme is involved in cleavage of photo-assimilate sucrose into glucose and fructose which is degraded in a very specific spatial and temporal pattern as part of the developmental program and is thus creating steep glucose gradients within the cotyledons upon seed development. Hence, sugars trigger seed developmental processes through a signal transduction network (Wobus and Weber, 1999). They also provide substrates for nucleic acid synthesis and for lignin, polyphenol and amino acid synthesis and glycolysis. Sucrose is the major transport form of carbon and is central to the export from the source and the import to the sinks (Meyer et al., 2007). Metabolites of the TCA

cycle, such as succinate and fumarate are highly ranked. Also highly ranked is myo-inositol. Other metabolites, such as glycerol-3-phosphate play a major role in membrane/phospholipid biosynthesis. Other highly-ranked metabolites are the amino acids alanine, isoleucine, leucine, methionine, serine, phenylalanine, proline, aspartate, tryptophan, tyrosine, valine, as well as the sugar alcohols myo-inositol and galactinol, and the fatty acids palmitate and linoleate and hydroxybutyrate.

Our data display the occurrence of both positive and negative correlations between metabolites and seed quality phenotypes. These findings corroborate the ideas of Meyer *et al.*, (2007) who found known metabolites displaying a negative correlation to the biomass. These metabolites are the aforementioned intermediates of central metabolic pathways including sucrose, glucose and the TCA cycle member's succinate and fumarate, as well as the amino acids alanine, isoleucine, leucine, methionine, serine, phenylalanine, proline, aspartate, tryptophan, tyrosine and valine. Although we have found both positively and negatively correlating metabolites amongst different seed quality phenotypes, the majority of the positively-correlated metabolites is a substantial fraction of metabolites related to stress responses, such as alanine, isoleucine, leucine, methionine, serine, phenylalanine, proline and valine, as well as some unknown metabolites. Thus, a link between the metabolites ranking high in the CCA and seed quality phenotypes is plausible because central metabolism and stress responses are of the utmost importance to seed germination, and thus, to seed quality. The observed scenario depicts the fact that positively-correlated metabolites could be an attribute to plant defense against abiotic and biotic stresses. Thus, these results suggest that higher concentrations of these metabolites coincide with better-armed plants. Another possible explanation for this is that positively correlated

metabolites are positive signals regulating plant growth and the contrary would be true for negatively-correlated metabolites.

Network Evaluation

An overview of the metabolite network was obtained by extracting all significant trait-trait correlations. Graph clustering using DPClus yielded densely-connected metabolites on the metabolomic correlation networks. KEGG enrichment analysis, used to assess the statistical significance of the detected clusters, demonstrated specific differences in the cluster in the enriched pathways.We postulate that the assigned KEGG pathways for each cluster reflect differences in underlying genetic properties of biochemical regulation of stage specific pathways. The largest cluster was 'Glyoxylate and dicarboxylate metabolism'. This cluster contained metabolites associated with the biosynthetic pathways of carbohydrates from fatty acids or precursors which enter the system as acetyl-coenzyme A. Its crucial enzymes are isocitrate lyase and malate synthase and they a relationship with several other metabolic processes: glycine, serine, and threonine, purine metabolism, carbon fixation, ascorbate and aldarate metabolism, nitrogen metabolism, pyruvate metabolism and the citrate cycle. Although there were no significant enriched KEGG pathways in clusters 3 and 5 (Figure 4.6), these clusters may represent the extensive coordination among biosynthetic pathways involved in fatty acids biosynthesis in tomato seeds. We followed an approach that may yield new insights into the organization of metabolites in the functional pathways of a given organism (Fukushima et al., 2011). Taken together, our observations demonstrate that variations in the topology of correlation networks reflect at least partially-known biochemical pathways in tomato (Camacho et al., 2005; Steuer, 2006; Fukushima et al., 2011). Our findings are in

agreement with Fukushima *et al.,* (2011) showing that graph clustering can be used to gather metabolites belonging to the metabolic pathways that change in response to different regulations. It is therefore likely correlation network is of more relevance than network similarity or proximity (Müller-Linow et al., 2007). Other studies support the idea of graph clustering approaches (Freeman et al., 2007; Fukushima et al., 2009; Fukushima et al., 2011). The aforementioned approaches have been applied effectively to gene co-expression networks for extracting functional, densely-connected genes. The present findings are consistent with a previous study which showed that the approach is also effective for metabolomic correlations (Fukushima et al., 2011).

Our study shows that dry and imbibed stages are associated with programmed metabolic switches during tomato seed development. Specific sets of metabolic components, distributed across the metabolic network, are synthesized during seed development according to need and possible utilization of certain metabolites. These biomarkers can be used concomitantly to predict increases in the flux of specific metabolites throughout the course of germination. The combined analysis of germination phenotypes and metabolite profiles provides a strong indication for the hypothesis that metabolic composition is related to germination phenotypes and thus to seed quality. The possibility of predicting germination traits on the basis of the metabolic signature presents a precedence for the use of metabolite profiles as biomarkers with high predictive power and could potentially revolutionize breeding for seed quality. The current findings contribute to our understanding of topology in the metabolic correlation network. They may add to the development of novel data-mining methods and to the discovery of biomarkers in order to understand complex molecular and genetic mechanisms underlying seed

quality. There are few systematic comparisons of metabolic correlations among data from time series under optimal and stress conditions (Byrne et al., 1998; Müller-Linow et al., 2007; Szymanski et al., 2009; Fukushima et al., 2011). Although we cannot demonstrate that metabolites are mapping to the same individual genes as seed quality QTLs, we can identify genetic regions that coordinate groups of metabolites and contain plausible candidate genes. Phenotypes mapping to the same locus may be co-regulated by the same genes. With our definition of "phenotype" now including metabolites and physiological traits we can begin to formulate relationships between these phenotypes and genetic regions. The groups of metabolites that are correlated or that co-map to physiological traits in our F_8 population may offer an insight into metabolic pathways that are causal or reactive to seed quality phenotypes. Our approach contributes to the generation of new testable hypotheses and may expand our fundamental understanding of metabolic behavior affected by genetic and/or environmental perturbations. The application of the GGG model allowed us to study the genetic basis of natural variation as well as environmental perturbations, i.e. differences between dry and imbibed seed profiles. In particular, utilization of this model in a natural population allowed us to study the plasticity of molecular networks by dissecting the genetic, as well as gene-by-environment, interaction components. Furthermore, it also detected how QTL effects may differ across different environments. The results of this study showed that heritable differences in genetic and environmental plasticity can be explored on a genome-wide scale. The uniqueness of this study presents a number of important implications for future practice for the characterization of unknown gene function(s) and helps in the high-throughput screening of metabolic phenotypes (Albinsky et al., 2010).

Acknowledgments
This work was supported by the Technology Foundation STW (R.K., L.W., W.L.) and by the Higher Education Commission, Pakistan (N.K.).

Supporting information can be downloaded from
www.wageningenseedlab.nl/thesis/rhkazmi/chapter4

Supplemental table 4.1. List of Metabolites analyzed and distribution among the two parents ('Pimp': *S. pimpinellifolium*: and 'MM': *S. lycopersicum*) of the studied RIL population

Supplemental table 4.2. Principle components and ANOVA p-values expressing the statistical significance of the changes in primary metabolites content across Dry and 6H tomato seed and discriminative metabolites of PCA loadings and significant changes in the metabolite levels of Dry and 6H

Supplemental table 4.3. Spearman R_s values and associated P Values for all pairwise correlations for known and all 160 metabolites

Supplemental table 4.4. QTL analysis results for metabolic traits

Supplemental table 4.5A. List of significantly correlated metabolites resulting from pairwise correlations (ordered by correlation)

Supplemental table 4.5B. List of all relevant metabolites determined by the correlation between them and the canonical variate (ordered by absolute correlation) and ranked according to the strength of the canonical correlation

Supplemental table 4.6. Resulting clusters detected by the DPClus algorithm. For details, see the manual of DPClus http://kanaya.naist.jp/DPClus/ webcite.

Supplemental table 4.7. List of KEGG pathways used in this study.

Supplemental Figure 4.1. Significant differences between *S. lycopersicum* and *S. pimpinellifolium* based on dry seeds. Metabolites above or below the grey box are more or less abundant, respectively, in *S. pimpinellifolium*. (B, C) Differences of metabolite concentration between dry seeds and 6h imbibed seeds for *S. lycopersicum* and *S. pimpinellifolium,* respectively. Metabolites above or below the grey box are more or less abundant, respectively, in 6h imbibed seeds. Metabolites are sorted on the x-axis from high abundance on the left to low abundance on the right. Metabolites indicated with a name are significantly different. P-values were calculated with a student T-test (n=3).

Supplemental Figure 4.2A Histogram of metabolite detection in the 101 RILs dependent upon detection in the MM and/or Pimp parental genotypes. The axis presents the number of metabolites found in a given number of RILs (on the axis). The axis separates the metabolites into four detection classes dependent upon whether the metabolites were found in the Money 'MM' and /or Pimp parents. ND means that the metabolite was detected in the given parental genotype. There were 90 metabolites detected in both parents, 7 metabolites detected in only Money 'MM', 10 metabolites detected in only Pimp and 53 metabolites detected in neither parent for dry seeds while 76 metabolites detected in both parents, 22 metabolites detected in only Money 'MM', 3 metabolites detected in only Pimp and 59 metabolites detected in neither parent for 6h imbibed seeds.

Supplemental Figure 4.2B Positive and negative transgressive segregation. Shown are distributions of glycerol-3-phosphate (A) and Hydroxylamine (B) within *S. lycopersicum* × *S. pimpinellifolium* RILs as examples of positive and negative transgressive segregation, respectively. Labels 'MM' and 'Pimp' show the average accumulation of these metabolites in the parental genotypes.

Supplemental Figure 4.3. PCA by removing four metabolites (threonate, pentonic acid and galactinol) in dry/6h imbibed using seeds datasets including 66 metabolites. See also the legend to Figure 1.

Supplemental Figure 4.4. Heat map of correlations between all 160 metabolites. Each square represents the Spearman correlation coefficient

between the metabolic phenotypes of the column with that of the row. Metabolic phenotype order is determined as in hierarchical clustering using the distance function 1-correlation.

Supplemental Figure 4.5A. A- Genomic locations of Genetic mQTLs identified for all 160 metabolites. Tomato chromosomes are identified by roman numerals (1–12), with centimorgans ascending from left to right; chromosomes are separated by white lines. Colored cells indicate QTL significant. Significant thresholds were defined with permutation analysis (n=1000, $p<0.01$) by randomizing the genotypes over each metabolite and was set to LOD >3 accordingly. The LOD color scale is indicated, showing blue and light blue when the *Solanum pimpinellifolium* ('Pimp') allele, and yellow and red when the *Solanum lycopersicum* (Money 'MM') allele, at that marker results in an elevated level of metabolic phenotype. QTL positions, LOD scores, effects are provided in Supplemental Table S4.4.

Supplemental Figure 4.5B. B- Genomic locations of G X E mQTLs identified for all 160 metabolites accumulation. Tomato chromosomes are identified by roman numerals (1–12), with centimorgans ascending from left to right; chromosomes are separated by white lines. Colored cells indicate QTL significant. Significant thresholds were defined with permutation analysis (n=1000, $p<0.01$) by randomizing the genotypes over each metabolite and was set to LOD >3 accordingly. The LOD color scale is indicated, showing blue and light blue when the *Solanum pimpinellifolium* ('Pimp') allele, and yellow and red when the *Solanum lycopersicum* (Money 'MM') allele, at that marker results in an elevated level of metabolic phenotype.QTL positions, LOD scores, effects are provided in Supplemental Table S4.4.

Supplemental Figure 4.6. Overview of distribution of overlapping metabolic and phenotypic QTLs. See legend to Figure S4.5A and S4.5B for description.

References:

Albinsky D, Kusano M, Higuchi M, Hayashi N, Kobayashi M, Fukushima A, Mori M, Ichikawa T, Matsui K, Kuroda H (2010) Metabolomic screening applied to rice FOX Arabidopsis lines leads to the identification of a gene-changing nitrogen metabolism. Molecular plant **3**: 125-142

Alexa A, Rahnenführer J, Lengauer T (2006) Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics 22: 1600-1607

Altaf-Ul-Amin M, Shinbo Y, Mihara K, Kurokawa K, Kanaya S (2006) Development and implementation of an algorithm for detection of protein complexes in large interaction networks. Bmc Bioinformatics **7**

Baena-González E, Sheen J (2008) Convergent energy and stress signaling. Trends in Plant Science 13: 474-482

Beaudoin F, Napier J (2004) 8 Biosynthesis and compartmentation of triacylglycerol in higher plants. Lipid Metabolism and Membrane Biogenesis: 33-55

Bewley JD, Black M (1994) Seeds: physiology of development and germination. Springer Us

Breitling R, Armengaud P, Amtmann A, Herzyk P (2004) Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. FEBS letters **573**: 83-92

Breitling R, Li Y, Tesson BM, Fu J, Wu C, Wiltshire T, Gerrits A, Bystrykh LV, De Haan G, Su AI (2008) Genetical genomics: spotlight on QTL hotspots. Plos Genetics **4:** e1000232

Buscher JM, Czernik D, Ewald JC, Sauer U, Zamboni N (2009) Cross-Platform Comparison of Methods for Quantitative Metabolomics of Primary Metabolism. Analytical Chemistry **81**: 2135-2143

Byrne P, McMullen M, Wiseman B, Snook M, Musket T, Theuri J, Widstrom N, Coe E (1998) Maize silk maysin concentration and corn earworm antibiosis: QTLs and genetic mechanisms. Crop science **38**: 461-471

Camacho D, de la Fuente A, Mendes P (2005) The origin of correlations in metabolomics data. Metabolomics **1**: 53-63

Carrari F, Baxter C, Usadel B, Urbanczyk-Wochniak E, Zanor MI, Nunes-Nesi A, Nikiforova V, Centero D, Ratzka A, Pauly M (2006) Integrated analysis of metabolite and transcript levels reveals the metabolic shifts that underlie tomato fruit development and highlight regulatory aspects of metabolic network behavior. Plant Physiology **142**: 1380

Carreno-Quintero N, Acharjee A, Maliepaard C, Bachem C, Mumm R, Bouwmeester H, Visser R, Keurentjes J (2012) Untargeted metabolic quantitative trait loci (mQTL) analyses reveal a relationship between primary metabolism and potato tuber quality. Plant Physiology

Chan EKF, Rowe HC, Hansen BG, Kliebenstein DJ (2010) The Complex Genetic Architecture of the Metabolome. Plos Genetics **6**

Chang CC, Beevers H (1968) Biogenesis of oxalate in plant tissues. Plant Physiology 43: 1821-&

Csardi G, Nepusz T (2006) The igraph software package for complex network research. InterJournal Complex Systems **1695**

Davies DD, Asker H (1983) Synthesis of oxalic acid by enzymes from lettuce leaves. Plant Physiology **72**: 134

Dixon RA, Harrison MJ, Lamb CJ (1994) Early events in the activation of plant defense responses. Annual Review of Phytopathology 32: 479-501

Eastmond PJ (2004) Glycerol-insensitive Arabidopsis mutants: gli1 seedlings lack glycerol kinase, accumulate glycerol and are more resistant to abiotic stress. The Plant Journal **37**: 617-625

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

Freeman TC, Goldovsky L, Brosch M, Van Dongen S, Mazière P, Grocock RJ, Freilich S, Thornton J, Enright AJ (2007) Construction, visualisation, and clustering of transcription networks from microarray expression data. PLoS computational biology **3**: e206

Fukushima A, Kanaya S, Arita M (2009) Characterizing gene coexpression modules in Oryza sativa based on a graph-clustering approach. Plant Biotechnol 26: 485-493

Fukushima A, Kusano M, Redestig H, Arita M, Saito K (2011) Metabolomic correlation-network modules in Arabidopsis based on a graph-clustering approach. Bmc Systems Biology **5**

Galili G (2011) The aspartate-family pathway of plants: Linking production of essential amino acids with energy and stress regulation. Plant Signaling & Behavior 6: 192

Garg AK, Kim JK, Owens TG, Ranwala AP, Do Choi Y, Kochian LV, Wu RJ (2002) Trehalose accumulation in rice plants confers high tolerance levels to different abiotic stresses. Proceedings of the National Academy of Sciences of the United States of America **99**: 15898-15903

Hurkman WJ, Tanaka CK (1996) Effect of salt stress on germin gene expression in barley roots. Plant Physiology 110: 971-977

Ilarslan H, Palmer RG, Imsande J, Horner HT (1997) Quantitative determination of calcium oxalate and oxalate in developing seeds of soybean (Leguminosae). American Journal of Botany **84**: 1042-1046

Jansen RC, Nap JP (2001) Genetical genomics: the added value from segregation. Trends in Genetics 17: 388-391

Jez JM, Cahoon RE (2004) Kinetic mechanism of glutathione synthetase from Arabidopsis thaliana. Journal of Biological Chemistry 279: 42726-42731

Kaplan F, Kopka J, Haskell DW, Zhao W, Schiller KC, Gatzke N, Sung DY, Guy CL (2004) Exploring the temperature-stress metabolome of Arabidopsis. Plant Physiology **136**: 4159

Kazmi RH, Khan N, Willems LAJ, van Heusden AW, Ligterink W, Hilhorst HWM (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 **5**: 929-951

Keurentjes J, Fu J, De Vos C, Lommen A, Hall R, Bino R, van der Plas L, Jansen R, Vreugdenhil D, Koornneef M (2006) The genetics of plant metabolism. Nature genetics **38**: 842-849

Keurentjes J, Sulpice R (2009) The role of natural variation in dissecting genetic regulation of primary metabolism. Plant Signaling & Behavior **4**: 244

Keurentjes J, Sulpice R, Gibon Y, Steinhauser M, 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

Keurentjes JJB, Bentsink L, Alonso-Blanco C, Hanhart CJ, Vries HBD, Effgen S, Vreugdenhil D, Koornneef M (2007) Development of a near-isogenic line population of Arabidopsis thaliana and comparison of mapping power with a recombinant inbred line population. Genetics **175**: 891-905

Keurentjes JJB, Fu J, Terpstra IR, Garcia JM, Van Den Ackerveken G, Snoek LB, Peeters AJM, Vreugdenhil D, Koornneef M, Jansen RC (2007) Regulatory network construction in Arabidopsis by using genome-wide gene expression quantitative trait loci. Proceedings of the National Academy of Sciences 104: 1708

Kliebenstein DJ (2009) Advancing Genetic Theory and Application by Metabolic Quantitative Trait Loci Analysis. Plant Cell **21**: 1637-1646

Koornneef M, Alonso-Blanco C, Vreugdenhil D (2004) Naturally occurring genetic variation in Arabidopsis thaliana. Annual Review of Plant Biology **55**: 141-172

Lane BG (1994) Oxalate, germin, and the extracellular-matrix of higher-plants. Faseb Journal **8**: 294-301

Li, Y., M. A. Swertz, G. Vera, J. Fu, R. Breitling, et al. (2009). DesignGG: An R-package and web tool for the optimal design of genetical genomics experiments. BMC Bioinformatics 10.

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

Libert B, Franceschi VR (1987) Oxalate in crop plants. Journal of Agricultural and Food Chemistry 35: 926-938

Lisec J, Meyer RC, Steinfath M, Redestig H, Becher M, Witucka-Wall H, Fiehn O, Torjek O, Selbig J, Altmann T, Willmitzer L (2008) Identification of metabolic and biomass QTL in Arabidopsis

thaliana in a parallel analysis of RIL and IL populations. Plant Journal **53**: 960-972

Lisec J, Steinfath M, Meyer RC, Selbig J, Melchinger AE, Willmitzer L, Altmann T (2009) Identification of heterotic metabolite QTL in Arabidopsis thaliana RIL and IL populations. Plant Journal **59**: 777-788

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

Mackay TFC, Stone EA, Ayroles JF (2009) The genetics of quantitative traits: challenges and prospects. Nature Reviews Genetics **10**: 565-577

Magrath R, Herron C, Giamoustaris A, Mithen R (1993) The inheritance of aliphatic glucosinolates in Brassica napus. Plant breeding **111**: 55-72

Meng PH, Raynaud C, Tcherkez G, Blanchet S, Massoud K, Domenichini S, Henry Y, Soubigou-Taconnat L, Lelarge-Trouverie C, Saindrenan P, Renou JP, Bergounioux C (2009) Crosstalks between Myo-Inositol Metabolism, Programmed Cell Death and Basal Immunity in Arabidopsis. PLoS One **4**

Meyer RC, Steinfath M, Lisec J, Becher M, Witucka-Wall H, Torjek O, Fiehn O, Eckardt A, Willmitzer L, Selbig J, Altmann T (2007) The metabolic signature related to high plant growth rate in Arabidopsis thaliana. Proceedings of the National Academy of Sciences of the United States of America **104**: 4759-4764

Millerd A, Morton RK, Wells JRE (1963) Oxalic acid synthesis in shoots of oxalis pes-caprae - precursors of glycollic acid and glyoxylic acid. Biochemical Journal 88: 276-&

Millerd A, Morton RK, Wells JRE (1963) Oxalic acid synthesis in shoots of oxalis pes-caprae (L.). Biochemical Journal **86**: 57-&

Mithen R, Magrath R (1992) Glucosinolates and resistance to Leptosphaeria maculans in wild and cultivated Brassica species. Plant breeding **108**: 60-68

Müller-Linow M, Weckwerth W, Hütt MT (2007) Consistency analysis of metabolic correlation networks. Bmc Systems Biology **1**: 44

Ogawa H, Gomi T, Konishi K, Date T, Nakashima H, Nose K, Matsuda Y, Peraino C, Pitot HC, Fujioka M (1989) Human-liver serine dehydratase - cdna cloning and sequence homology with hydroxyamino acid dehydratases from other sources. Journal of Biological Chemistry **264**: 15818-15823

Phillips PC (2008) Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. Nature Reviews Genetics **9**: 855-867

Price GD, Badger MR, Woodger FJ, Long BM (2008) Advances in understanding the cyanobacterial CO2-concentrating-mechanism (CCM): functional components, Ci transporters, diversity, genetic regulation and prospects for engineering into plants. Journal of Experimental Botany **59**: 1441

Prinzenberg AE, Barbier H, Salt DE, Stich B, Reymond M (2010) Relationships between Growth, Growth Response to Nutrient Supply, and Ion Content Using a Recombinant Inbred Line Population in Arabidopsis. Plant Physiology **154**: 1361-1371

Rajasekaran LR, Aspinall D, Jones GP, Paleg LG (2001) Stress metabolism. IX. Effects of salt stress on trigonelline accumulation in tomato. Canadian Journal of Plant Science **81**: 487-498

Roessner U, Wagner C, Kopka J, Trethewey RN, Willmitzer L (2000) Simultaneous analysis of metabolites in potato tuber by gas chromatography–mass spectrometry. The Plant Journal **23**: 131-142

Rowe HC, Hansen BG, Halkier BA, Kliebenstein DJ (2008) Biochemical networks and epistasis shape the Arabidopsis thaliana metabolome. Plant Cell 20: 1199-1216

Saito K (1996) Formation of L-ascorbic acid and oxalic acid from D-glucosone in Lemna minor. Phytochemistry **41**: 145-149

Sanchez DH, Pieckenstain FL, Szymanski J, Erban A, Bromke M, Hannah MA, Kraemer U, Kopka J, Udvardi MK, Provart N (2011) Comparative Functional Genomics of Salt Stress in Related Model and Cultivated Plants Identifies and Overcomes Limitations to Translational Genomics. PLoS One **6**: e17094

Schauer N, Semel Y, Balbo I, Steinfath M, Repsilber D, Selbig J, Pleban T, Zamir D, Fernie AR (2008) Mode of inheritance of primary metabolic traits in tomato. Plant Cell **20**: 509-523

Schauer N, Semel Y, Roessner U, Gur A, Balbo I, Carrari F, Pleban T, Perez-Melis A, Bruedigam C, Kopka J, Willmitzer L, Zamir D, Fernie AR (2006) Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. Nature Biotechnology **24**: 447-454

Shen W, Wei Y, Dauk M, Zheng Z, Zou J (2003) Identification of a mitochondrial glycerol-3-phosphate dehydrogenase from Arabidopsis thaliana: evidence for a mitochondrial glycerol-3-phosphate shuttle in plants. FEBS letters **536**: 92-96

Skogerson K, Harrigan GG, Reynolds TL, Halls SC, Ruebelt M, Iandolino A, Pandravada A, Glenn KC, Fiehn O (2010) Impact of Genetics and Environment on the Metabolite Composition of Maize Grain. Journal of Agricultural and Food Chemistry **58**: 3600-3610

Stacklies W, Redestig H, Scholz M, Walther D, Selbig J (2007) pcaMethods - a bioconductor package providing PCA methods for incomplete data. Bioinformatics **23**: 1164-1167

Steuer R (2006) Review: on the analysis and interpretation of correlations in metabolomic data. Briefings in bioinformatics **7**: 151-158

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

Sulpice R, Trenkamp S, Steinfath M, Usadel B, Gibon Y, Witucka-Wall H, Pyl ET, Tschoep H, Steinhauser MC, Guenther M, Hoehne M, Rohwer JM, Altmann T, Fernie AR, Stitt M (2010) Network Analysis of Enzyme Activities and Metabolite Levels and Their Relationship to Biomass in a Large Panel of Arabidopsis Accessions. Plant Cell **22**: 2872-2893

Szymanski J, Jozefczuk S, Nikoloski Z, Selbig J, Nikiforova V, Catchpole G, Willmitzer L (2009) Stability of metabolic correlations under changing environmental conditions in Escherichia coli–a systems approach. PLoS One **4**: e7441

Tikunov Y, Laptenok S, Hall R, Bovy A, de Vos RCH (2011) MSClust: a tool for unsupervised mass spectra extraction of chromatography-mass spectrometry ion-wise aligned data. Metabolomics: **1-5**

Tolbert N (1981) Metabolic pathways in peroxisomes and glyoxysomes. Annual review of biochemistry **50**: 133-157

Toubiana D, Semel Y, Tohge T, Beleggia R, Cattivelli L, Rosental L, Nikoloski Z, Zamir D, Fernie AR, 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

Trethewey RN, Geigenberger P, Riedel K, Hajirezaei MR, Sonnewald U, Stitt M, Riesmeier JW, Willmitzer L (1998) Combined expression of glucokinase and invertase in potato tubers leads to a dramatic reduction in starch accumulation and a stimulation of glycolysis. The Plant Journal **15**: 109-118

Ullmann P, Gondet L, Potier S, Bach TJ (1996) Cloning of Arabidopsis thaliana glutathione synthetase (GSH2) by functional complementation of a yeast gsh2 mutant. European Journal of Biochemistry **236**: 662-669

Voorrips RE, Verkerke W, Finkers R, Jongerius R, Kanne J (2000) Inheritance of taste components in tomato. Acta Physiologiae Plantarum 22: 259-261

Wang CL, Oliver DJ (1996) Cloning of the cDNA and genomic clones for glutathione synthetase from Arabidopsis thaliana and complementation of a gsh2 mutant in fission yeast. Plant Molecular Biology **31**: 1093-1104

Wang W, Ballatori N (1998) Endogenous glutathione conjugates: Occurrence and biological functions. Pharmacological Reviews **50**: 335-355

West MAL, Kim K, Kliebenstein DJ, Van Leeuwen H, Michelmore RW, Doerge R, Clair DAS (2007) Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in Arabidopsis. Genetics **175**: 1441-1450

Wobus U, Weber H (1999) Sugars as signal molecules in plant seed development. Biological Chemistry **380**: 937-944

Chapter 5
DISSECTION OF THE COMPLEX PHENOTYPES OF SEED QUALITY ON TOMATO CHROMOSOMES 6 AND 8: HIFS (HETEROGENEOUS INBRED FAMILIES) CONFIRMATION

Kazmi RH, Willems LAJ, Ligterink W, Hilhorst HWM

In preparation

Abstract

Seed quality is a complex trait and therefore the genetics of seed phenotypes is poorly understood. Numerous quantitative trait loci (QTLs) with moderate to large phenotypic effects that influence tomato seed quality attributes were identified previously from a *Solanum lycopersicum (*cv. Moneymaker, 'MM') x *Solanum pimpenellifolium* (G1.1554, 'Pimp') recombinant inbred line (RIL) population. Here we report on the successful confirmation of some of those complex loci by recapitulating the original linked phenotypes by an analysis of near-isogenic lines (NILs) developed from the RIL population that differ at the respective QTLs. To confirm and validate QTLs a specific type of near-isogenic line, so called heterogeneous inbred families (HIFs) that segregate for a genomic region of interest, were used. HIFs differ for markers linked to a QTL of interest extracted from segregating families. The application of this procedure is described for QTLs associated with several germination traits in tomato. A population of 83 RILs was screened with two CAPS markers for the identification of HIFs that were associated with seed germination traits. Two segregating families

were identified for each marker from linkage groups 6 and 8. The progeny of these HIFs were characterized for the segregation of seed germination phenotypes for markers flanking the QTLs. NILs derived from each HIF had significantly differences in germination traits, confirming a number of loci that influence seed quality in tomato. The use of HIFs accelerates subsequent QTL fine-mapping and may thus lead sooner to the identification of genes potentially involved in the control of different linked physiological seed processes.

Introduction

Deciphering the genetic basis of natural variation in quantitative traits is a challenging task. The identification of causal sequence variants demands for QTL confirmation and their molecular characterization. Although substantial progress has been made in cloning genes causal for QTLs and reducing some of them to Quantitative Trait Nucleotides (QTNs), QTL characterization remains a challenging task. Difficulties arise as QTL mapping produces large genetic intervals and QTLs of large effect can split into multiple QTLs, explaining only a small proportion of the total variance (Balasubramanian et al., 2009). This makes it a formidable task to determine the causal genes. Furthermore, phenotypic effects of individual QTLs are often even more complicated by the phenotypic variability resulting from segregation of other loci that influence the same trait (Loudet et al., 2005). However, near-isogenic lines (NILs) that differ for markers flanking the QTL may offer a solution (Robertson et al., 1988; Kaeppler et al., 1993; Kooke et al., 2012). Analysis of the NILs allows the dissection of QTLs into smaller intervals as they differ for overlapping regions of the genome, as indicated by QTL analysis (Tuinstra et al., 1997).

Hence, NILs can be used to map QTLs to smaller genomic intervals, as well as for phenotypic confirmation of specific QTLs. Thus, NILs offer a common genetic background in which direct comparison of two lines can be used to evaluate the phenotype conditioned by a QTL (Dorweiler et al., 1993; Touzet et al., 1995; Tanksley and Nelson, 1996; Eduardo et al., 2005). Despite the long success of NILs in narrowing down QTL intervals, their use has been limited mainly because of the substantial effort needed to develop appropriate genetic material. Considerable effort has been made to address the problems associated with developing NILs that differ at individual QTLs. Marker-assisted selection over one or two further generations in advanced backcross populations (AB-QTL) for identified QTLs can lead to NILs contrasting at the specific QTLs (Tanksley and Nelson, 1996). Thus, a combination of a QTL mapping experiment with subsequent studies to confirm and evaluate specific QTLs in NILs may result in smaller QTL intervals (Tanksley et al., 1996). Alternatively, NILs contrasting at QTLs can be developed by selection within heterogeneous inbred families (HIFs) (Tuinstra et al., 1997). HIFs are a set of lines derived from RILs that are genetically similar but have not reached complete homozygosity yet and segregate for those loci that were heterozygous. Molecular markers can be used to screen a population of HIFs derived from different inbreds to identify families that segregate for a specific region of the genome (Tuinstra et al., 1997). This approach can be used to develop a series of NILs that contrast for a specific genomic region. This HIF concept is effective because one does not first have to create the NILs, which requires several generations of backcrossing and marker-assisted selection (MAS).

Quantitative genetics has enjoyed a renaissance in the past decade. QTL analysis has yielded a long list of genomic regions in different

agriculturally important species that are responsible for a wide range of phenotypic traits. The causal genes for several of these QTLs have been identified (Ligterink et al., 2012; Rajjou et al., 2012). The recent identification of various seed quality phenotypes in tomato and Arabidopsis have increased the demand for molecular-genetic dissection of these traits (Joosen et al., 2012; Kazmi et al., 2012; Khan et al., 2012; Rajjou et al., 2012). Previously, we have used a recombinant inbred line (RIL) population generated from *Solanum lycopersicum (*cv. Moneymaker, 'MM') and *Solanum pimpenellifolium* (G1.1554, 'Pimp') to identify QTLs for seed quality phenotypes (Kazmi et al., 2012). QTL clusters were identified across the 12 tomato chromosomes that influence seed germination in both control- and stress conditions. These QTL clusters are very useful for breeding purposes as selection for rapid germination in one condition may lead to progeny rapidly germinating under other conditions. QTL mapping with RIL populations results in a very limited resolution (Tuinstra et al., 1997; Keurentjes et al., 2007) and reducing the size of the QTL clusters will be very useful for both gene identification and for breeding purposes. The aim of this study was to evaluate and confirm seed quality QTLs on chromosomes 6 and 8. In this study HIFs of tomato RILs that differ at QTLs associated with different seed quality traits were used to test for QTL effects on chromosomes 6 and 8 for germination potential. Taking advantage of the residual heterozygosity and the large size of the RIL population, we were able to confirm QTLs for onset and rate of germination (t_{10}^{-1}, t_{50}^{-1}, respectively) as well as area under the curve (AUC) in HIFs under control-, salt-, osmotic-, high-temperature and oxidative stress conditions.

Material and Methods

QTL mapping

For QTL analysis the *S. lycopersicum* (cv. Moneymaker, 'MM')× *S. pimpinellifolium* (G1.1554, 'Pimp') RIL population was used, which has been exhaustively mined for a large number of seed quality QTLs (Kazmi et al., 2012). The mapping software MapQTL®5.0 (Van Ooijen and Maliepaard, 2003) was used for identifying QTL positions in the genome for a given trait. A multiple QTL mapping model (MQM) was used to identify potential QTLs (Jansen et al., 1995) as implemented in MapQTL®5.0.

Development, growth conditions and seed collection of heterogeneous inbred families (HIFs)

HIF lines were derived from residual heterozygosity remaining in some of the F_8 RILs at the loci of interest. For each of these lines, 20 plants were individually genotyped at the segregating markers and two homozygous plants for each of the parental alleles (MM or Pimp) were selected and selfed to produce seeds for further phenotypic analysis.

Two HIFs segregating for each QTL were identified in the population. A set of NILs was selected from each of the segregating HIFs for chromosomes 6 and 8. The HIF233 (chromosome 6) set consisted of a set of NILs of which 6 lines had a Pimp background, 5 lines a MM background and 11 lines a heterozygous background. In the case of HIF271 (chromosome 8) we could select 2 lines with Pimp-, 4 lines with MM- and 2 lines with a heterozygous background. These lines were genotyped with CAPS markers across the tomato genome to determine the average heterogeneity of each HIF. The segregation of other markers flanking the target QTL was evaluated to estimate the size of the heterogeneous region, differentiating the lines in each HIF. These HIFs

were grown under controlled conditions in the greenhouse facilities at Wageningen University, the Netherlands. The day and night temperatures were maintained at 25 and 15 °C, respectively, with 16 h light and 8 h dark (long-day conditions). The cleaned seeds were dried for 3 d at 20 °C and were stored in a cool, dry storage room (13 °C and 30% RH) in paper bags.

HIFs genotyping

DNA was extracted using a method previously described (Cheung et al., 1993). PCR was performed on this DNA in a Bio-Rad S1000^{TM} Thermocycler. The PCR reactions were conducted for 4 min at 95 °C, followed by 35 cycles of 20 sec denaturation step at 95 °C, primer annealing step at varying temperatures; 30 sec at 55 °C, 1 min at 72 °C, and 10 min extension step at 72 °C. PCR products were run on a 1.5 % agarose gel stained with Gel Red and the genotype of the plant was assessed. The primers used for the PCRs were: 42712640 (physical position of marker in bp), Forward – TTTCTCAGTCGAGCTCAT, Reverse – ATACCCATAGACTTGCTG on chromosome 6; whereas for chromosome 8 15684096 (physical position of marker in bp), Forward – CTCAAATCAAGAACGCTGAC, Reverse – GTTAGTTGTGGGTTGCAT were used.

Phenotypic characterization

Germination assay

To characterize differences in different seed quality phenotypes, germination assays were performed in triplicate with seeds of the parents and the HIFs, which were sown under aseptic conditions on germination trays (21 × 15 cm; DBP Plastics NV, Antwerpen, Belgium, http://www.dbp.be) containing 15 mL water (non-stress condition) or NaCl,

polyethylene glycol (PEG) or H_2O_2 (stress conditions) and one layer of white filter paper (20.2 × 14.3 cm white blotter paper; Allpaper BV, Zevenaar, The Netherlands, http://www.allpaper.nl). Each germination tray contained two lines with 45 seeds of each line. Germination trays were placed in a completely randomized design with three replications per sample. A maximum of 12 trays were piled up with two empty trays on both the top and the bottom end of the stack, with 15 mL water and two layers of white filter paper, to prevent unequal evaporation. The trays were covered with tightly fitting lids and the whole pile was wrapped in a closed transparent plastic bag and incubated at 4 °C for 3 d for stratification. Subsequently, the bags where placed randomly in an incubator at 25 °C in the dark (type 5042; Seed Processing Holland), except for brief intervals when germination was counted under laboratory (fluorescent) lighting. Germination responses were scored visually as radicle protrusion at eight hourly intervals for 10 consecutive days during the period of most rapid germination, and at longer subsequent intervals, until no additional germination was observed.

Salt, osmotic, high-temperature and oxidative stress

Salt, osmotic and oxidative stress tolerance treatments were applied in germination trays with 15 mL of the corresponding solution on a piece of filter paper to confirm the seed quality traits on chromosomes 6 and 8. Response to salt stress was estimated by germinating HIF233 seeds in NaCl. Osmotic potentials were established through aqueous solutions of polyethylene glycol (PEG 8000; Sigma, St Louis, MO, USA) measured in megapascal (MPa). Specific concentrations of NaCl (-0.3MPa) and PEG 8000 (-0.3MPa) were determined with the Solute Potential and Molar-Molal-g Solute/g Water Interconversion (SPMM) program (Michel and Radcliffe,

1995). HIF233 was subjected to suboptimal temperature regimes in order to test their response to temperature stress. Germination was monitored during incubation for 10 d at 35°C to test for high-temperature stress response. Tolerance to hydrogen peroxide was estimated by germinating HIF271 seeds on filter paper saturated with a solution of 300 mM H_2O_2.

Calculation of G_{max}, t_{10}^{-1}, t_{50}^{-1}, U_{7525}^{-1}, AUC and estimation of means

The curve-fitter module of the Germinator package was used to analyze different parameters of the cumulative germination curves (Joosen et al., 2010). Parental lines and the HIF populations were subjected to different germination conditions, and maximum germination (G_{max}, %), the onset of germination [t_{10}^{-1}; reciprocal of time to 10% of viable seeds to germinate (h^{-1})], the rate of germination [t_{50}^{-1}; reciprocal of time to 50% of the viable seeds to germinate (h^{-1})], MGR = mean germination rate, which is reciprocal of the mean germination time (MGT^{-1}), uniformity [U_{7525}^{-1}, reciprocal of time interval between 75 and 25% viable seeds to germinate; h^{-1}] and area under the germination curve [AUC; the integration of the fitted curve between t = 0 and a user-defined end point; 200h] were determined as described previously (Kazmi et al., 2012).

Results

Identification of QTLs associated with seed phenotypes

QTL analysis was performed on the basis of the established marker linkage map of the *S. lycopersicum (*cv. Moneymaker, 'MM')× *S. pimpinellifolium* (G1.155, 'Pimp') RIL population, which contains 865 SNP markers (Kazmi et al., 2012). Numerous QTLs were identified with moderate to large phenotypic effects for 42 tomato seed quality traits. Highly significant levels of overlapping QTLs between phenotypic traits was

also established by permutation tests conducted on all −1LOD QTL intervals. Seven QTL clusters positioned on chromosomes 1, 2, 4, 6, 8, 9 and 12 were identified as affecting several seed germination traits with overlapping proportions ranging from 62.5 to 100% at −1LOD (Kazmi et al., 2012).

Development of near-isogenic lines

HIFs were constructed to validate the results of the QTL analysis using the previously described methodology (Tuinstra et al., 1997; Alonso-Blanco and Koornneef, 2000; Loudet et al., 2005; Joosen et al., 2012). NILs were developed by screening heterogeneous inbred families (HIFs) to identify families that were heterogeneous for the SNP marker most tightly associated with each QTL (Figure 5.1). The RIL is chosen that is still heterozygous at the QTL, then selfed and genotyped (F_8 plants) so that each homozygous genotype at the region of interest can be identified and studied in detail. HIFs are not to be compared with the reference parental genotype but with one another within the descendants (family) of the chosen RIL. In contrast with 'conventional' NILs, the genetic background is not homogeneous, but a mix of both parental genomes, since these lines originate from one RIL of the population. Numerous loci were investigated in order to find candidate RILs for the construction of the HIFs to validate the QTL hotspots mapped on different chromosomes. We could not find HIFs to validate all the detected QTLs clusters, but after screening the F_8 RILs we were able to find candidate RILs for the construction of the HIFs for QTLs on chromosome 6 originating from 952233 (HIF233) and chromosome 8 originating from 952271 (HIF271). These HIFs were screened with CAPS marker 43582592 on chromosome 6 whereas CAPS marker 15684096 was used on chromosome 8. HIF233 is shown to

segregate for marker 442712640 on chromosome 6 whereas HIF271 is shown to segregate for CAPS marker 15684096 on chromosome 8. An example of selection and respective genotyping is shown for chromosome 6 in Figure 5.1.

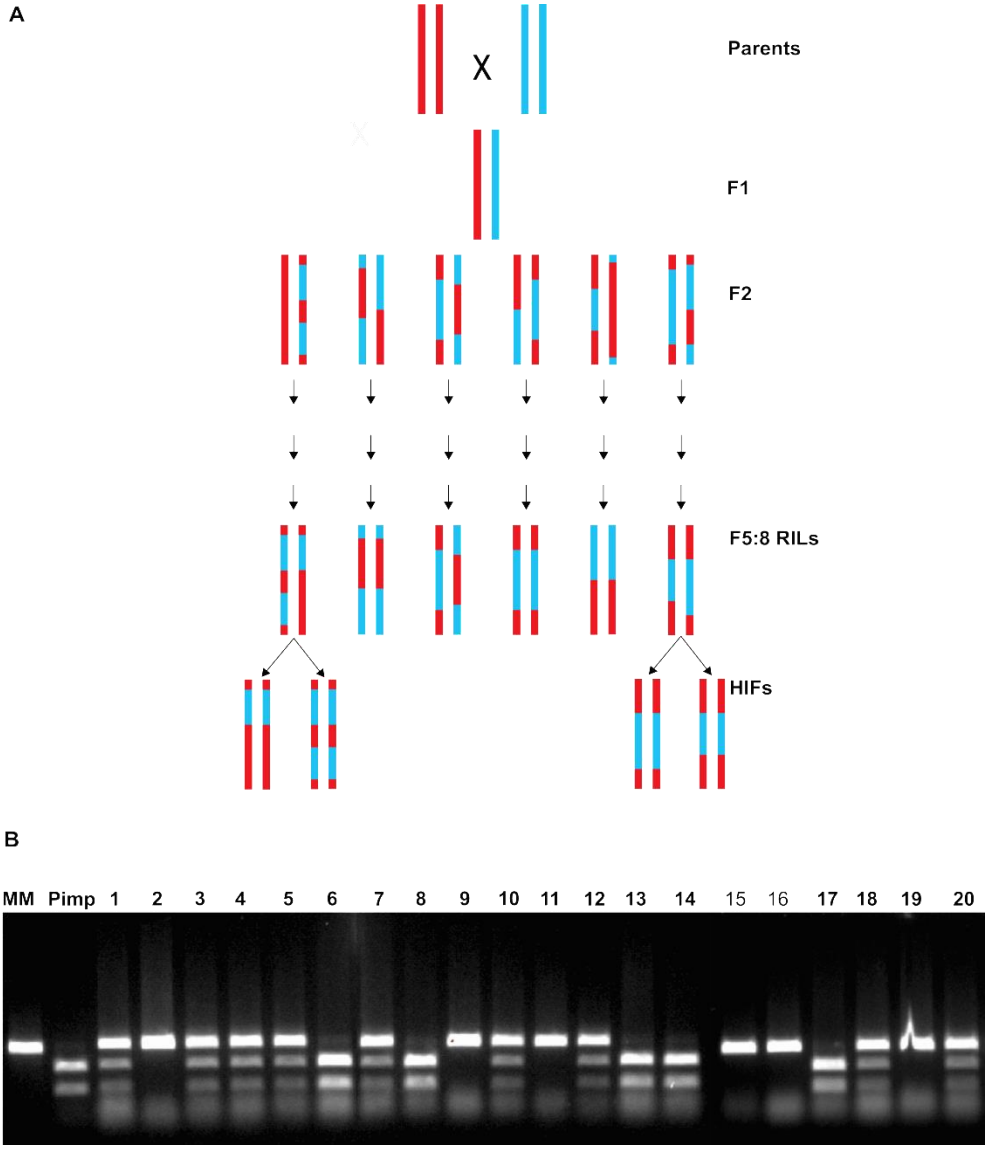


Figure 5.1 (A) The production of heterogeneous inbred families (HIFs) from a non-fixed recombinant inbred population. The graphical genotype of

individual plants is depicted for a representative single pair of chromosomes. Near isogenic lines (NILs) can be produced from inbred lines that are not entirely homozygous obtained by continuing selfing until the F_7 generation (single-seed descent). Analysis of RILs with molecular markers around the QTL of interest allows selection of heterozygous lines in this region. The phenotypic examination of further selfing progeny in combination with further genotyping enables the selection of NILs that are in a mixed heterogeneous genetic background. Thus, pairs of NILs differing at marker loci associated with quantitative trait loci can be selected to obtain heterogeneous families (Tuinstra et al., 1997; Alonso-Blanco and Koornneef, 2000; Loudet et al., 2002). (B) Example of screening of progeny lines to select heterogeneous inbred families using the CAPS marker 42712640 on chromosome 6. Progeny in HIF233 are segregating for the marker; where parent MM and Pimp are shown on the left. The HIFs with money background show one band, for Pimp background two bands and heterozygous HIFs are depicted with three bands.

Confirmation of QTLs in near-isogenic lines

Confirmation of the QTLs on chromosome 6 for germination under non-stress conditions

A number of significant occurrences of overlapping QTLs among G_{max}, t_{10}^{-1}, t_{50}^{-1}, U_{7525}^{-1} and AUC were observed after permutation tests among all QTL positions in previous QTL mapping studies for seed quality phenotypes (Kazmi et al., 2012). For instance, at the bottom of chromosome 6, the confidence intervals for a number of germination trait QTLs significantly overlapped with those detected across different environmental conditions. This interval harbors overlapping QTLs for different germination curve descriptors under normal and stress

environments. The germination phenotypes were calculated only for those traits in which a corresponding fraction (10, 50 and 75% or more) of seeds completed germination.The germination assays were performed under normal conditions to confirm the QTLs for different germination traits e.g. G_{max}, t_{10}^{-1}, t_{50}^{-1}, U_{7525}^{-1} and AUC under control as well as stress conditions. HIF233 showed a difference between MM and Pimp alleles in the region surrounding marker 42712640 for different germination traits. Figure 5.2 illustrates the genotypes and phenotypes of these lines, showing that t_{50}^{-1}, did segregate in the interval in HIF233.

Confirmation of the salt, osmotic and high-temperature stress QTLs on chromosome 6

Comparing the salt-, osmotic-, and high-temperature responses of HIF 233, the localization and the additive effects of the number of QTLs in the same interval related to different germination traits, as mentioned above, were confirmed (Figure 5.2). In HIF233, segregating for marker 42712640, lines carrying the MM allele had a significantly faster germination (higher t_{50}^{-1}) and higher AUC than lines with the Pimp allele under salt stress conditions. The MM allele at the 42712640 marker was also responsible for a higher osmotic tolerance than the Pimp allele for t_{50}^{-1} and AUC, as lines carrying MM alleles had a significantly faster germination (t_{50}^{-1}) and higher AUC than lines derived from Pimp. To investigate the specificity of the high-temperature stress germination QTL, we also evaluated the behavior of HIF233 seeds under high-temperature. Germination differences were found between lines with MM and Pimp backgrounds and, thus, we concluded that this locus is also involved in regulating G_{max}, t_{50}^{-1} and AUC specifically in conferring tolerance to high-temperature stress (Figure 5.2). The data demonstrate that the MM alleles

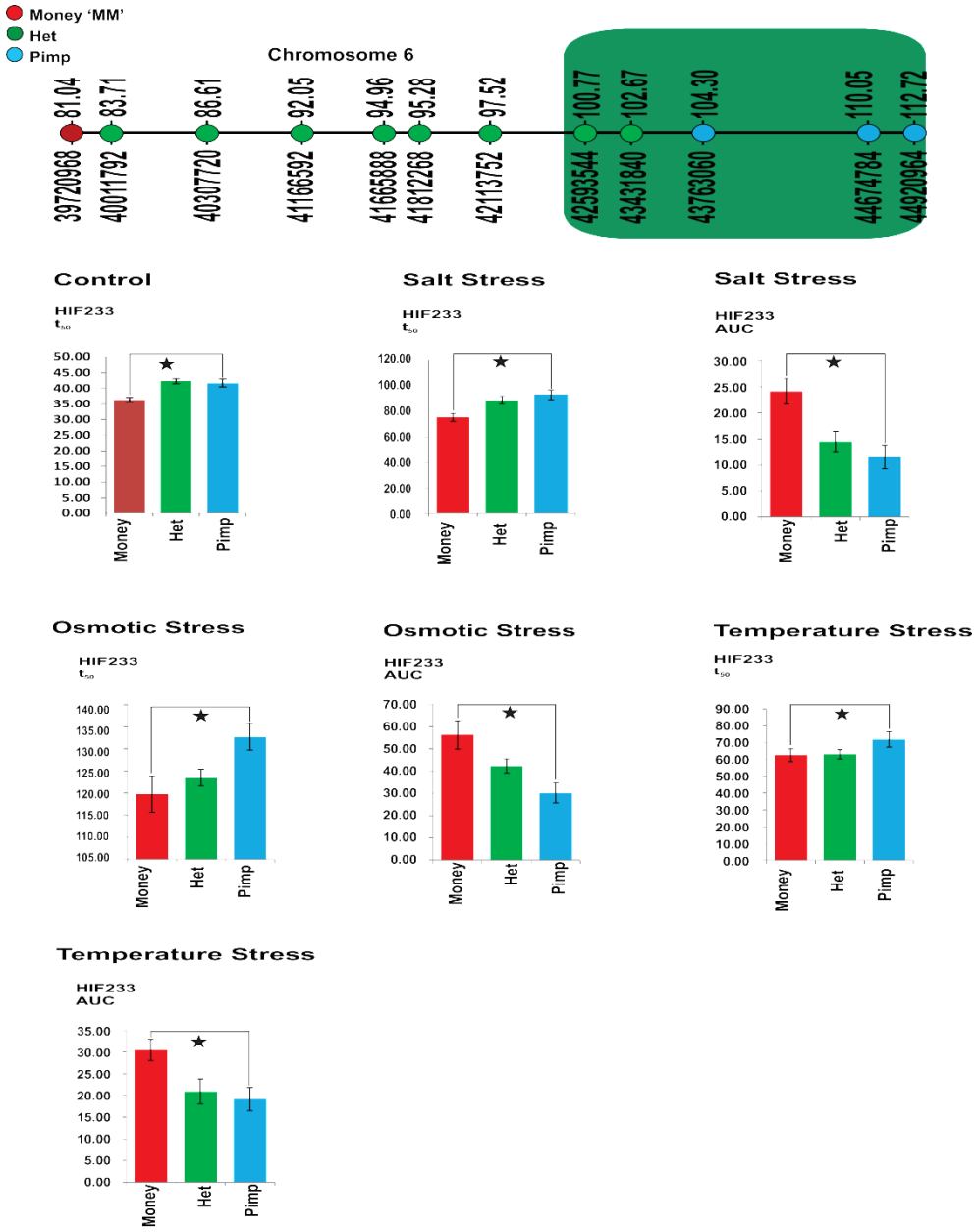


Figure 5.2. (A) Fine mapping and phenotype characterization of heterogeneous inbred families segregating for seed quality phenotypes. (A)

Genetic map of QTL on chromosome 6 for seed quality phenotypes. Circles represent molecular markers, with their corresponding names in megabases below and positions incentimorgan (cM) on top. The name of each marker on the tomato linkage map corresponds to the position on the tomato genome sequence version SL2.40 (http://solgenomics.net/organism/solanum_lycopersicum/genome). The colored boxed area represents the 1-LOD confidence interval of the QTLs. HIF233 is heterozygous for part of the confidence interval of the QTLs. (B) Phenotypes for the progeny of HIF233. Rate of germination (t_{50}) and area under the curve (AUC) were measured in each HIF line under control and stress (salt, osmotic, high-temperature) conditions. Asterisks indicate significant differences between MM and Pimp alleles.

at these markers confer tolerance to salt-, osmotic-, and high-temperature stresses during germination, corroborating the observations in the original mapping population previously used to map these QTLs (Kazmi et al., 2012).

Confirmation of the oxidative stress QTLs on chromosome 8

To confirm the localization and the additive effects of the QTL locus on the middle region of chromosome 8, polymorphic HIFs originating from 952271 were selected in the region of interest (Figure 5.3). The germination traits of the progeny lines derived from HIF271 were measured to determine the association between germination traits. HIF271 confirmed the presence of the previously identified QTLs for number of germination traits (G_{max}, t_{10}^{-1} and AUC) at the 15684096 marker because the germination traits carrying the MM alleles in HIF271 at this marker showed

a higher tolerance to oxidative stress during germination than those carrying Pimp alleles at the same marker (Figure 5.3).

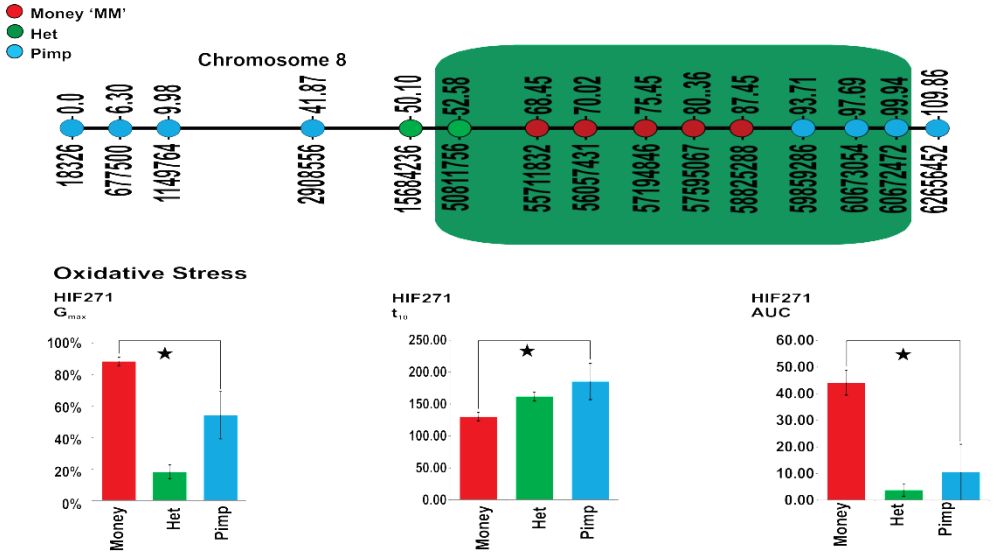


Figure 5.3. Fine mapping and phenotype characterization of heterogeneous inbred families segregating for seed quality phenotypes on chromosome 8. (A) Genetic map of seed quality phenotypes region on chromosome 8. Circles represent molecular markers, with their corresponding names in megabases below and positions in centimorgan (cM) on top. The name of each marker on the tomato linkage map corresponds to the position on the tomato genome sequence version SL2.40 (http://solgenomics.net/organism/solanum_lycopersicum/genome). The colored boxed area represents the 1-LOD confidence interval of the QTLs. (B) Phenotypes for HIF233 lines. Maximum germination (G_{max}), onset/rate of germination (t_{10}) and area under the curve (AUC) were

measured in each HIF line under oxidative stress conditions. Asterisks indicate significant differences between MM and Pimp alleles.

Discussion

Immortal populations are a powerful tool to study quantitative traits. The analysis of the Quantitative Trait Loci (QTLs) obtained in this way, remain a formidable task, as this produces large genetic intervals and molecular characterization of the allelic variation requires that the respective genes are cloned. In order to delineate the QTL intervals for the identification of causal genes, there is need to incorporate additional genome-wide information, which is becoming increasingly accessible. High-precision mapping leading to fine mapping and cloning of gene(s) responsible for complex trait variation usually requires the construction of NILs that differ only at a small region at the QTL of interest (Glazier et al., 2002). Normally NILs are produced by introgressing one parent allele at the QTL region in the other parents' background, and this process requires several backcrosses. In present study NILs were produced by taking advantage of residual heterozygosity which is left in the F_8 generation of the used RIL population, to find lines that still segregate only at the region of the QTL of interest. This type of NILs are called heterozygous inbred families (HIFs) (Tuinstra et al., 1997; Alonso-Blanco and Koornneef, 2000; Loudet et al., 2002; Loudet et al., 2005; Joosen et al., 2012).

QTLs affecting germination phenotypes under non-stress (control) and stress conditions at the bottom of chromosome 6 were confirmed using HIFs. Similarly, our investigation of the material described in this report has also revealed the differing genetic control of seed germination in the MM and Pimp accessions for the QTLs for germination under oxidative stress on chromosome 8. QTLs co-locating for G_{max}, t_{10}^{-1} and AUC under oxidative

stress at a locus on chromosome 8 were characterized by testing HIF271. We confirmed that the variation observed in the previous phenotyping experiments in the RILs is relevant and repeatable. In a previous study by Kazmi *et al.*, (2012), QTLs for different seed phenotypes were often found clustered across the 12 tomato chromosomes, in particular at the bottom of chromosome 6. Seeds of HIF233 carrying MM alleles were more tolerant than those carrying Pimp alleles for germination under salt-, osmotic- and high-temperature stresses. HIF analyses confirmed the co-locating QTLs for t_{50}^{-1} and AUC under normal-, salt-, osmotic-, and high-temperature stress conditions. Confirmation of these QTLs controlling seed germination under control and stressed conditions strengthens our previous findings of the presence of genetic relationships between the ability of seeds to germinate rapidly under different environments. Interestingly, inside the QTL cluster on chromosome 6, the allelic effects are in the same direction, which strengthens the possibility that one gene accounts for all the co-localized QTLs. Hence, it could conceivably be hypothesized that the QTLs validated at the bottom of chromosome 6 across the various germination environments are candidates that can be used in marker-assisted selection (MAS) or gene cloning.

This combination of findings provides some support for the conceptual premise for the further fine mapping and cloning of the gene(s) responsible for complex trait variation, which usually requires the construction of NILs in order to determine the exact genetic relationships among these traits. Hence, isolation, characterization and comparison of functional gene analyses of these loci can be combined with a candidate gene approach by looking at all the genes in the confidence interval. The examination of QTL for seed germination traits in tomato indicated that HIF analysis could identify NILs that were useful for confirming linkage between

markers and QTL, for fine mapping of QTL, and for evaluating the phenotype associated with specific QTLs. The RIL population as well as the HIFs showed variation for different germination traits under control as well as salt-, osmotic-, high-temperature and oxidative stress conditions. The strategy to be adopted for further fine-mapping is to detect recombination in the interval by screening descendants of heterozygous plants of each HIF and to use these, once fixed, as new HIFs with smaller candidate regions. Furthermore, the comprehensive analyses of the genetic architecture of complex traits may require the consideration of multiple populations that represent a larger sample of the standing intraspecific variation and thus provide a framework for comparative analyses (Ahn and Tanksley, 1993). The existence of a number of QTLs identified for tomato remains to be investigated by exploiting the heterozygosity using HIFs that would allow the unequivocal identification of QTL regions across the tomato genome.

The present study provides additional evidence in revealing the differing genetic control of seed quality traits. HIF analysis provided an efficient method to develop NILs in tomato in order to dissect the genetic architecture behind seed germination phenotypes subjected to different contrasting environments (control-, vs. stressed). Different germination descriptors are particularly interesting when studied in interaction with stress environments, e.g. under salt-, osmotic-, high-temperature and oxidative stress. HIF analysis can be used to generate NILs in a range of recombinant genetic backgrounds and are useful for identifying a genetic background in which the phenotype of a QTL is clearly expressed. This should facilitate detailed phenotypic characterization of individual QTLs, particularly QTLs that are strongly influenced by the environment and

genetic background. The cloning of the underlying genes will provide a basis to optimize seed germination efficiency and thereby seed quality.

References:

Ahn S, Tanksley S (1993) Comparative linkage maps of the rice and maize genomes. Proceedings of the National Academy of Sciences **90:** 7980-7984

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

Balasubramanian S, Schwartz C, Singh A, Warthmann N, Kim MC, Maloof JN, Loudet O, Trainer GT, Dabi T, Borevitz JO (2009) QTL mapping in new Arabidopsis thaliana advanced intercross-recombinant inbred lines. PLoS One **4:** e4318

Chardon F, Virlon B, Moreau L, Falque M, Joets J, Decousset L, Murigneux A, Charcosset A (2004) Genetic architecture of flowering time in maize as inferred from quantitative trait loci meta-analysis and synteny conservation with the rice genome. Genetics **168:** 2169-2185

Cheung W, Hubert N, Landry B (1993) A simple and rapid DNA microextraction method for plant, animal, and insect suitable for RAPD and other PCR analyses. Genome Research **3:** 69-70

Dorweiler J, Stec A, Kermicle J, Doebley J (1993) Teosinte glume architecture 1: a genetic locus controlling a key step in maize evolution. Science **262:** 233-235

Eduardo I, Arus P, Monforte AJ (2005) Development of a genomic library of near isogenic lines (NILs) in melon (Cucumis melo L.) from the exotic accession PI161375. TAG Theoretical and Applied Genetics **112:** 139-148

Glazier AM, Nadeau JH, Aitman TJ (2002) Finding genes that underlie complex traits. Science **298:** 2345-2349

Jansen R, Ooijen JW, Stam P, Lister C, Dean C (1995) Genotype-by-environment interaction in genetic mapping of multiple quantitative trait loci. TAG Theoretical and Applied Genetics **91:** 33-37

Joosen RVL, Arends D, Willems LAJ, Ligterink W, Jansen RC, Hilhorst HWM (2012) Visualizing the genetic landscape of Arabidopsis seed performance. Plant Physiology **158:** 570-589

Joosen RVL, Kodde J, Willems LAJ, Ligterink W, Van Der Plas LHW, Hilhorst HWM (2010) germinator: a software package for high-

throughput scoring and curve fitting of Arabidopsis seed germination. The Plant Journal **62:** 148-159

Kaeppler S, Phillips R, Kim T (1993) Use of near-isogenic lines derived by backcrossing or selfing to map qualitative traits. TAG Theoretical and Applied Genetics **87:** 233-237

Kazmi RH, Khan N, Willems LEOAJ, Van Heusden AW, Ligterink W, Hilhorst HWM (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

Keurentjes JJB, Bentsink L, Alonso-Blanco C, Hanhart CJ, Blankestijn-De Vries H, Effgen S, Vreugdenhil D, Koornneef M (2007) Development of a near-isogenic line population of Arabidopsis thaliana and comparison of mapping power with a recombinant inbred line population. Genetics **175:** 891-905

Khan N, Kazmi RH, Willems LAJ, van Heusden AW, Ligterink W, Hilhorst HWM (2012) Exploring the Natural Variation for Seedling Traits and Their Link with Seed Dimensions in Tomato. PLoS One **7:** e43991

Kooke R, Wijnker E, Keurentjes J (2012) Backcross populations and near isogenic lines. Methods in molecular biology (Clifton, NJ) **871:** 3

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

Loudet O, Chaillou S, Camilleri C, Bouchez D, Daniel-Vedele F (2002) Bay-0× Shahdara recombinant inbred line population: a powerful tool for the genetic dissection of complex traits in Arabidopsis. TAG Theoretical and Applied Genetics **104:** 1173-1184

Loudet O, Gaudon V, Trubuil A, Daniel-Vedele F (2005) Quantitative trait loci controlling root growth and architecture in Arabidopsis thaliana confirmed by heterogeneous inbred family. TAG Theoretical and Applied Genetics **110:** 742-753

Michel BE, Radcliffe D (1995) A computer program relating solute potential to solution composition for five solutes. Agronomy Journal **87:** 126-130

Rajjou L, Duval M, Gallardo K, Catusse J, Bally J, Job C, Job D (2012) Seed germination and vigor. Annual review of plant biology **63:** 507-533

Robertson D, Victor B, Helfman G, Schultz E, Warner R, Searcy W, Pleszczynska W, Hansell R, Lili A, Hollman S (1988) Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature **335:** 20

Tanksley S, Grandillo S, Fulton T, Zamir D, Eshed Y, Petiard V, Lopez J, Beck-Bunn T (1996) Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative L. pimpinellifolium. TAG Theoretical and Applied Genetics **92:** 213-224

Tanksley S, Nelson J (1996) Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. TAG Theoretical and Applied Genetics **92:** 191-203

Touzet P, Winkler R, Helentjaris T (1995) Combined genetic and physiological analysis of a locus contributing to quantitative variation. TAG Theoretical and Applied Genetics **91:** 200-205

Tuinstra M, Ejeta G, Goldsbrough P (1997) Heterogeneous inbred family (HIF) analysis: a method for developing near-isogenic lines that differ at quantitative trait loci. TAG Theoretical and Applied Genetics **95:** 1005-1011

Van Ooijen JW, Maliepaard C (2003) MapQTL®, Version 5.0:Software for the Calculation of QTL Positions on Genetic Maps.

Chapter 6
GENERAL DISCUSSION; CREATING SYSTEM-LEVEL MODELS OF TOMATO SEED QUALITY

Kazmi RH, Ligterink W, Hilhorst HWM

1.1 Seed quality

Seed is a living product that must be grown, harvested and processed correctly in order to maximize its viability and subsequent crop productivity. Seed quality is the ability of seeds to germinate under a wide variety of environmental conditions and to develop into healthy seedlings. Seed quality is determined by several factors including genetic and physical purity, mechanical damage and physiological attributes, such as viability, germination, dormancy, vigor and uniformity (Dickson, 1980; Hilhorst and Toorop, 1997; Hilhorst, 2007; Hilhorst, 2010). The physiological condition of seeds during development and maturation has a strong effect on ultimate seed quality. Several environmental factors, such as temperature, humidity, light and nutrients during the seed filling and maturation stages, seed treatments (harvesting and processing) and accumulated damage all influence seed quality (Ouyang et al., 2002; Spano et al., 2007). Thus, seed quality is a complex trait governed by interactions between the genome and the environment (Koornneef et al., 2002) and, therefore, seed quality can be challenged over the entire seed production chain.

1.2 Natural variation for complex trait of tomato seed quality

Seed quality is a complex trait and quantitative in nature, yet very little is understood, at the molecular/genetic level, about the final physiological state of seed quality; in other words how seed quality is

established. This thesis seeks to remedy these problems by giving attention to the genetics of seed quality in order to understand underlying genetic mechanisms regulating different seed processes and their relationship to seed and seedling quality. Exploitation of the natural variation present in *S. lycopersicum* x *S. pimpinellifolium* Recombinant Inbred Line (RIL) population with regard to seed and seedling phenotypic states and rapid growth of molecular markers provide the opportunity to understand the physiological and genetic mechanisms governing seed quality. Thus, to relate and reason upon networks of bio-molecular and phenotypic variables inorder to identify novel alleles for tomato seed quality at various levels of organization were highly desirable. Thus, in response to this gap in knowledge and practice, we report on a comprehensive study that utilizes a multi-disciplinary approach to dissect the components of seed quality. This study was intended to elicit meaningful and significant biomarker-phenotype complexes, spanning multiple levels of refinement.

1.3 Genetic analysis of germination phenotypes

Primarily, as a classical first step in genetic analysis of seed and seedling phenotypes the natural variation present in a *S. lycopersicum* x *S. pimpinellifolium* recombinant inbred line (RIL) population was explored. The RIL population obtained from these accessions proved to be a powerful resource for the detection of seed quality Quantitative Trait Loci (QTLs). The QTL approach appears to be valuable not only in elucidating the genetics, but also the physiological background of seed quality phenotypes. Chapter 2 began by laying out the theoretical dimensions of the research, and looked at how the genetic variation that is present in the RIL population controls the regulation of various germination indices (Kazmi et al., 2012). The final germination of seeds is one of the qualitative attributes

of the germination process; it portrays the overall germination potential of crop species based on a binary answer: germinated or non-germinated. There is consensus as to the meaning, methods and calculation of germinability in time or at the end of the observations (Ranal and Santana, 2006). Although, final germination (G_{max}) is an important factor for estimating the expected seedling yield of a seed lot, which can be partly independent of other important germination characteristics, such as rate and onset of germination (t_{50}^{-1}/t_{10}^{-1}, MGR = mean germination rate), as well as uniformity (U_{7525}^{-1}) of germination. Thus, it was important to include the various aspects of cumulative germination in order to quantify the different seed quality traits under the various germination conditions. The quantification of the germination responses was simplified, as both the rate and the percentage of germination could be incorporated into the 'area under curve' (AUC). The analysis of germination was enriched by communicating the onset/rate and AUC values in addition to the final germination, hence measuring different aspects of the germination process. This study demonstrated the usefulness of these germination parameters in describing the extremes of pattern differences of seed germination. It has been shown that germination parameters are under strong genetic control (El-Kassaby, 1991). Therefore, analyzing different aspects of cumulative germination curves is an important phenotypic attribute of a seed lot and is of importance with respect to the consequences of genetic diversity for seed quality present in the *S. lycopersicum* × *S. pimpinellifolium* RIL population.

The QTLs were mapped to different genomic regions on the 12 chromosomes of tomato for different germination descriptors. The mapped QTLs showed a variable number of overlapping QTL clusters, which is quite understandable as they are descriptors of the same germination-time

curve. However, inspection of the QTLs affecting individual parameters across different chromosomes also revealed significant hot spots for one parameter but not for others. These results suggest that there are specific loci that affect certain germination characteristics, but not all. Furthermore, co-location of roughly two-thirds of the QTLs affecting germination traits across different stresses, highlights the positive relationship between seed quality phenotypes and different stress types. The study also corroborated previous QTL mapping studies of germination under salt, drought and cold stresses in tomato where 71% of the detected QTLs affected germination under two or more stresses, indicating that common factors are associated with different germination conditions. Seed germination under different stress conditions was genetically controlled, with additivity being the major genetic component. Significantly large genetic correlations between germination responses at different stress environments indicate that similar or identical genes might contributed to the germination response under these different stress conditions. Thus, selection for rapid germination at one stress condition would result in progeny with improved germination under diverse stress conditions. It has been shown that germination of tomato is genetically controlled and hence can be increased by selection (Dudley, 1993; Tanksley, 1993; Foolad et al., 2003; Foolad, 2007). Productive and sustainable crop growth necessitates growing plants in suboptimal environments with less input of precious resources. This study was intended to take a step forward towards better understanding and rapid improvement of abiotic stress tolerance in tomato, and to link physiological and underlying molecular mechanisms of seed quality. In terms of seed vigor, the rate and uniformity of germination is a sensitive indicator of a high-quality seed, and these attributes deteriorate more

quickly than final germination; they therefore represent a key component to seed quality.

1.4 Genetic analysis of seedling traits and their link with seed dimensions

Seedlings are living, perishable plants, profoundly impacted upon by genetic, physiological and environmental factors. Morphological assessments of seedlings give us information about the physical manifestation of the seedlings' physiological response to the growth environment (Mexal and Landis, 1990). The seedlings' ability for shoot penetration through the impeding soil of the seed bed is an essential attribute of vigor (Khan et al., 2012). The ability to germinate faster, followed by seedling growth, represents key phenotypic markers for seed vigor regulated by the genetic architecture of plant species (Bettey et al., 2000). Seedling quality and subsequent field performance can be influenced by various stress factors. Thus, a vigorous seed should possess the ability to establish healthy seedlings across various environments. Seedling shoot height, fresh and dry root and shoot weights are the most common measures used for growing and estimating the downward growth rate of roots and upward growth rate of shoots, as well as predicting seed vigor (Bettey et al., 2000; Epstein, 2004; Fita et al., 2008). There are many additional morphological parameters that can be assessed as well. No single factor has been shown to provide a perfect prediction of planting success, but each of them has been linked with seedling performance potential in some way. Measuring seedling quality can help to identify possible crop problems in order to make informed decisions for culturing, lifting, storing and planting. One of the more significant aspects of the present study was its emphasis on seed dimensions, such as seed size and

seed length, which has been ignored in previous studies. Although seed size, length and weight are interdependent traits, this study showed significant differences in the total number of the mapped QTLs for these seed dimensions. The strong association among seed weight and seedling traits under nutrient deficient conditions also corroborates the notion that larger seeds are better able to establish or survive as seedlings in a variety of environments, including nutrient shortage (Lee and Fenner, 1989; Jurado and Westoby, 1992).

The current findings add substantially to our understanding of the quantification of underground parts, as studies on roots are lagging behind those of shoots (Epstein, 2004). For tomato, no relevant information is available on root growth-related traits, nor has any proper study on seedling growth been published. Root systems are important to plant survival as they perform the crucial task of providing water, nutrients and physical support to the plant. The length of the primary/main root and the number of lateral roots are important components of root architecture, and play a key role in determining the success of a plant in a particular environment (Malamy and Benfey, 1997). So far, however, there has been no discussion about the genetic analysis of seedling traits in tomato and, to the best of our knowledge, this is the first genetic analysis, adding to a growing body of literature on root architecture. The current study, described in chapter 3, found strong relationships between different seed/seedling dimensions and root architecture, cementing the argument that larger food reserves in large-sized seeds help in establishing a more extensive root system. Evidently, an efficient root system ultimately aids in the acquisition of nutrients and the uptake of water from lower layers of soil under low-nutrient and low-moisture conditions, thereby playing an important role in the utilization of nutrients from the soil (Baker, 1972;

Zhang et al., 1999). The RIL population used in this study showed genetic variation of the analyzed seed and seedling traits, as a number of hot spots regulating these traits were found across the tomato genome.

The overlapping QTL clusters were evident along the tomato genome for the seed dimensions and seedling traits. A strong relationship between seed traits (seed weight, size and vigor) on the initial downward growth of the root system has been reported in the literature, in addition to its effect on the upward growth of seedlings (Baker, 1972; Jurado and Westoby, 1992). Several tomato genotypes with heavier seeds produced heavier seedlings, compared to genotypes with small seeds (Nieuwhof et al., 1989). Positive effects of heavy seeds, as well as higher quantities of reserve food in larger seeds as, compared to small seeds (Hilhorst et al., unpublished results), could be due to common genetic mechanisms controlling these traits. This study has been unable to demonstrate significant correlations between seed size or seed weight and seed performance, such as rate and uniformity of germination or maximum germination percentage (Kazmi et al., 2012), as was found in other species (Fenner, 1991). It seems possible that seed size is beneficial to the establishment of seedlings, but there appears to be no consistent link between seed size and germination characteristics. Furthermore, it was also evident that germination performance and seed size are controlled by different independent genetic loci (Kazmi et al., 2012).

1.5 Generalized genetical genomics and molecular phenotyping

Recent advances in genome sequencing and high-throughput phenotyping technologies have enabled plant biologists to study

quantitative genetics and the association of genetic variation of the phenotypes by elucidating the regulatory processes underlying complex traits at a multi-dimensional level (transcriptomics, proteomics and metabolomics). Thecombination of metabolomics with quantitative genetics is at the heart of understanding biochemical phenotypes. In particular, rather than focusing on seed quality as a single phenotypic entity, we explored pathways and networks by which to better predict the consequences of genetic and metabolic variations on the complex trait of seed quality. Chapter 4 describes a systems-genetics approach in which genetic perturbations of seed quality were studied across multiple environments. A comprehensive understanding of biological systems requires studying them across multiple environments as the molecular networks largely depend on environmental cues (Li et al., 2008; Ruffel et al., 2010). Thus, this allows a crucial level of understanding to study the system's behavior by varying the experimental conditions. While previous solutions to this problem are promising, none have taken the method of generalised genetical genomics (GGG) that was used in the present study for metabolite abundance profiling. Using GC-TOF-MS, the genetic regulation of variation in the tomato primary polar seed metabolome of the tomato RIL population could be described. GGG uses environmental perturbations (different seed developmental stages, i.e. dry and imbibed seeds) in combination with the analysis of the genetic variation present in the RIL population, to study the change of metabolites over the multiple environments and to identify genotype-by-environment interactions; hence, it can be used to reveal the plasticity of molecular networks for seed quality traits in tomato. It is interesting to note that dry and 6h imbibed seeds were associated with programmed metabolic switches during tomato seed germination. Different metabolites were shown to be synthesised in

accordance with demand and possible utilization. The progression of seeds from the dry to the imbibed stage was associated with changes in levels of the majority of amino acids and their precursors - alcohols, sugars, organic acids and fatty acid compounds. This suggests that metabolism during the 6h seed stage has to render certain metabolites rapidly available to support metabolic recovery during imbibition. It seems possible that primary metabolites might be rapidly consumed to support the metabolic switch toward enhancing biosynthetic processes needed for early germination. Thus, these metabolic changes can act as markers to predict the increase in the flux of specific metabolites throughout the course of tomato seed germination.

Central to the entire discipline of metabolite profiling is the concept of pathways converging to common metabolites. These pathways have multiple controllers; just one genetic locus may not alter metabolite levels significantly and therefore may not be identified as a metabolic quantitative trait locus (mQTL). Nonetheless, in our F_8 population, we found significant linkage signals, including some that are quite strong, providing evidence for the co-regulation of biologically-related pathways. Metabolites were mapped to distinct genetic regions, much like mRNA transcripts (eQTL). Similar to distribution of the phenotypic QTLs along the genome, hotspots and empty regions were observed for metabolite abundance. The amino acid metabolites provided the most striking evidence of functional clustering. It is interesting to note that both the correlation matrix and the genetic linkage data showed that most of the amino acids group together as was also shown by Toubiana *et al.*, (2012) when studying a introgression lines (IL) population. The clustering of metabolites based on their relative concentration may be explained by the fact that different genetic regulators drive these metabolites and play a key role in assigning

them unique signatures even within a group of highly correlated metabolites. These clustering patterns of signature metabolites suggest that static metabolic profiling can be used as a marker for changes in flux through certain metabolic pathways. However, a number of mQTLs mapped to multiple positions, which indicates a complex regulation of the tomato seed metabolome.

The large-scale genetic analysis of metabolite abundance clearly shows its usefulness in predicting germination phenotypes and constructing correlation network modules. Multivariate statistics proved to be an important approach by which to predict the germination phenotypes from a large set of metabolites. In general, therefore, it seems that integrative biomarkers have a highly significant positive or negative correlation with germination parameters and capture much of the information present in the metabolite profile (Meyer et al., 2007; Lisec et al., 2008; Sulpice et al., 2010). Natural variation present for germination phenotypes coincides with characteristic combinatorial changes of metabolite levels, whereas individual metabolites may fluctuate largely, independently of alterations in germination. The question of the metabolic regulatory mechanism and the role of metabolites was further disentangled by having a complementary overview of the metabolite network obtained by extraction of all significant trait-trait correlations. KEGG enrichment analysis, used to assess the statistical significance of the detected clusters, demonstrated specific differences in the clusters in the enriched pathways, yielding new insights into the organization of metabolites in functional pathways (Fukushima et al., 2011). Taken together, our observations demonstrate that variations in the topology of correlation networks reflect at least partially-known, biochemical pathways in tomato (Camacho et al., 2005; Steuer, 2006; Fukushima et al., 2011).

1.6 Post QTL analyses and methods

QTL analysis remains a formidable task as it yields large genetic intervals whereas molecular characterization of the allelic variation requires that the respective genes are cloned. In order to delineate the QTL intervals for the identification of causal genes, it is necessary to incorporate additional genome-wide information, which is becoming increasingly accessible. High-precision mapping, leading to fine mapping and cloning of gene(s) responsible for complex trait variation, usually requires the construction of near isogenic lines (NILs) that differ only in a small region at the QTL of interest (Glazier et al., 2002). Chapter 5 describes the synthesis, construction, characterization and evaluation of NILs to validate QTL hot spots present at the bottom of chromosomes 6 and 8. In this investigation, the aim was achieved by taking advantage of residual heterozygosity left in the F_8 generation of our RIL population in order to find lines that still segregated only at the region of the QTL of interest. This type of NIL is called a 'heterogeneous inbred family' (HIF) (Tuinstra et al., 1997; Alonso-Blanco and Koornneef, 2000; Loudet et al., 2002; Loudet et al., 2005). The QTLs were confirmed using HIFs affecting germination phenotypes under non-stress (control) and stress conditions at the bottom of chromosome 6. QTLs for different seed phenotypes were often found clustered across the 12 tomato chromosomes and in particular at the bottom of chromosome 6. Confirmation of these QTLs controlling seed germination under control and stress conditions further strengthened our findings that the genetic variation observed in the previous phenotyping experiments in the RILs is relevant and repeatable, and the ability to germinate rapidly under different germination conditions is genetically regulated (Chapter 2).

QTL clusters validated by HIFs correspond to the QTL clusters

identified previously at the bottom of chromosome 6 and in the middle of chromosome 8 in the mapping population (Kazmi et al., 2012). A number of marker associated with QTLs confirmed by HIFs at the bottom of chromosome 6 across different germination conditions are candidates that can be used in marker-assisted selection (MAS) or help in gene cloning by fine mapping. Nonetheless, this locus operates as a common molecular component of the signaling network that controls germination under various stress conditions. Therefore, pleotropic effects of this locus for different seed germination traits support the genetic dissection of seed quality in order to facilitate a more strategic approach to breed for resultant better seed quality in tomato. These findings provide some support for the conceptual premise for the further fine mapping and cloning of the gene(s) responsible for complex trait variation, which usually requires the construction of NILs in order to determine the exact genetic relationships among these traits. The present study makes several contributions to dissecting the complex trait (s) of seed quality. Moreover, isolation, characterization and comparison of functional gene analyses of these loci can be combined with a candidate gene approach by looking at all the genes in the confidence interval. However, there are still a number of other genetic regions alongside of tomato genome previously identified, which remain that need to be confirmed by exploiting the heterozygosity (HIFs' identification) present in RILs that would allow the unequivocal recognition of QTL regions across tomato genome.

1.7 Integrative analysis: key aspects, issues, Limitations and solutions

The present study shows that there is extensive genetic variation present in phenotypes (germination, seed and seedling traits) and metabolite profiles. The combined datasets from several QTL mapping (phenotypic, metabolic) studies reflect the genetic changes responsible for seed quality differences among multiple RILs of the *Solanum lycopersicum* x *Solanum pimpinellifolium* population. Moreover, the application of high throughput metabolite profiling enabled the construction of metabolic networks, which were unlikely to be uncovered from targeted small-scale approaches. Taken together, a large number of QTLs were identified, yet the identification of causal genes at the nucleotide level (QTN), affecting the trait of interest remains a challenging task. Several limitations of QTL mapping need to be acknowledged as they produce large genetic intervals. In addition, large-effect QTLs can split into multiple QTLs, explaining only a small proportion of the total variance (Balasubramanian et al., 2009). Furthermore, phenotypic effects of individual QTLs are often even more complicated because of the phenotypic variability resulting from segregation of other loci that influence the same traits. We believe that genetical genomics allows the direct quantification of the link between genotypes and their genetic responses. An important assumption in a typical genetical genomics studies is that the gene/metabolite expression values are also affected by what is causing the differences among the traits. The most important limitation lies in the fact that the genetical genomics approach by which we attempted to narrow the gap between QTL and phenotype, was not full accomplished. Unfortunately, because of time constraints, we could not integrate the e(xpression) QTL data in this study. This has somewhat limited our current efforts to comprehend the

genetic regulation of the mechanisms underlying complex traits by integrating genotypic, phenotypic, gene expression and metabolic data. The potential and applications of integrative genetical genomics have been described in several studies (Morley et al., 2004; Kadarmideen et al., 2006; Keurentjes et al., 2008). However, the data from microarray profiling for the tomato RIL population is in the process of analysis and will be available for incorporation with phenotypic and metabolomic data sets in the near future to obtain a comprehensive picture at a system level.

1.8 Visions: systematic characterization of tomato seed quality

The exploratory findings in this study provide a new understanding of tomato seed quality at the system level. Genes underlying QTLs act through complex networks of transcriptional, protein and metabolic phenotypes, complicating the comprehension of the genetics behind complex traits and the downstream changes that they reflect. Most important is the challenge to understand the causative and correlative effects of genetic perturbations on these networks and their downstream effects on organismal phenotypes. These challenges should be met in the near future by applying modeling of how changes/perturbations in one gene will affect the expression of other genes and what effect those changes will have on the phenotypes. The dramatic increase in high-throughput technologies and their integrative analytical procedures are opening the way to molecular, physiological and genetic plasticity studies, triggered by environmental perturbations (Jansen and Nap, 2001; Li et al., 2006; Li et al., 2008). Returning to the question posed at the beginning of this study, it is now possible to state that systems genetics promise to integrate these layers of information in order to produce

directed biological networks that link molecular variants to organismal phenotypes. However, sophisticated statistical techniques (Bayesian networks, partial correlation analysis and empirical Bayes procedures) are needed to construct meaningful networks from natural genetic perturbations (Mackay et al., 2009). Disentangling causal relationships from consequential relationships is the key to reconstructing biological networks Keurentjes *et al*., (2007) used the power of systems genetics to understand the biological basis of variation for quantitative traits and the potential regulatory network construction by combining eQTL mapping and candidate gene selection in Arabidopsis.

Promising approaches and methods are being developed to investigate further genome-wide gene specificity and regulation through alternative splicing (Wang et al., 2008) and microRNAs (Yin et al., 2008), next generation sequencing (NGS) (Shendure and Ji, 2008), RNA-Seq (Wang et al., 2009). RNA-Seq is a new arrival in the era of technological revolutions for transcriptome profiling that uses deep-sequencing technologies. It enhances the precision of transcriptomic measurement in characterizing the eukaryote transcriptomes (Wang et al., 2009). The application of next generation sequencing (NGS) technologies to the genomics field could change the experimental design dramatically and significantly improve eQTL studies. Next generation sequencing enables the analysis of complete genomes without the need of a pre-known sequenced genome, although it does require a close reference genome with which it can be compared. This technology opens up a great opportunity to study those organisms in which a genome has not been sequenced; and, unlike microarrays, it does not have to cope with sequence cross-hybridization problems nor with microarray experimental designs. In addition, the application of NGS also

allows a direct link between genotypes and phenotypes (Shendure and Ji, 2008).

1.9 Final remarks

Our results show integrative approaches that bridge different disciplines especially the integration of multiple layers of biological information from DNA to phenotype, with which to dissect the underlying molecular mechanism of seed quality in order to create system-level models of seed quality, which were notably lacking in the contemporary research in tomato crop. Therefore, novel combinatorial physiological, genetics and genomics approaches were used to create formalized knowledge collections based upon the information incumbent to physiological and genetic principles regulating seed quality. Furthermore, our results indicate that such systems-level models contain interrelated bio-molecular, physiological and metabolic markers capable of supporting hypothesis discovery and testing. This research had, as its aim, the identification of novel and knowledge-anchored biomarker-phenotype complexes. However, there is no single recipe for the dissection of complex traits nor for the selection of different datasets and their analysis. Clearly, we integrated different data sets, and a crucial step is managing, creating and analyzing the data, but ultimately the characterization and interpretation of complex traits is better obtained through integrating various methodologies. The extent and complexity of such studies make it impossible to perform these analyses by single individuals. Here we described a framework and procedures that can guide researchers through the complete process and expand the characterization of known QTLs responsible for a certain complex trait. This research has raised many questions with the need for further investigation. A number of possible

future studies using the same experimental set up are apparent, and the combining of data from several QTL mapping datasets (phenotypic, metabolic), most importantly data sets from eQTLs, and examining the genetic changes responsible for seed quality in the *Solanum lycopersicum* x *Solanum pimpinellifolium* population are highly recommended.

References:

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

Baker H (1972) Seed weight in relation to environmental conditions in California. Ecology **53:** 997-1010

Balasubramanian S, Schwartz C, Singh A, Warthmann N, Kim MC, Maloof JN, Loudet O, Trainer GT, Dabi T, Borevitz JO (2009) QTL mapping in new Arabidopsis thaliana advanced intercross-recombinant inbred lines. PLoS ONE **4:** e4318

Bettey M, Finch-Savage W, King G, Lynn J (2000) Quantitative genetic analysis of seed vigour and pre-emergence seedling growth traits in Brassica oleracea. New Phytologist **148:** 277-286

Camacho D, de la Fuente A, Mendes P (2005) The origin of correlations in metabolomics data. Metabolomics **1:** 53-63

Dickson M (1980) Genetic aspects of seed quality. Horticultural Science **15:** 771-774

Dudley JW (1993) Molecular markers in plant improvement - manipulation of genes affecting quantitative traits. Crop Science **33:** 660-668

El-Kassaby YA (1991) Genetic variation within and among conifer populations: review and evaluation of methods. *In* Fineschi S, Malvolti M.E., Cannata F, H H.H., eds, Biochemical markers in the population genetics of forest trees SPB Academic Publishing bv, The Hague, pp 61-76

Epstein E (2004) Plant biologists need to get back to their roots. Nature **430:** 829

Fenner M (1991) The effects of the parent environment on seed germinability. Seed Science Research **1:** 75-84

Fita A, Pico B, Monforte A, Nuez F (2008) Genetics of Root System Architecture Using Near-isogenic Lines of Melon. Journal of the American Society for Horticultural Science **133:** 448-458

Foolad MR (2007) Molecular Mapping, Marker-Assisted Selection And MAP-Based Cloning In Tomato. *In* Rajeev K. Varshney, R Tuberosa, eds, Genomics-Assisted Crop Improvement, Volume 2: Genomics Applications in Crops, pp 307-356

Foolad MR, Subbiah P, Kramer C, Hargrave G, Lin GY (2003) Genetic relationships among cold, salt and drought tolerance during seed germination in an interspecific cross of tomato. Euphytica **130:** 199-206

Fukushima A, Kusano M, Redestig H, Arita M, Saito K (2011) Metabolomic correlation-network modules in Arabidopsis based on a graph-clustering approach. Bmc Systems Biology **5**

Fukushima A, Kusano M, Redestig H, Arita M, Saito K (2011) Metabolomic correlation-network modules in Arabidopsis based on a graph-clustering approach. Bmc Systems Biology **5:** 1

Glazier AM, Nadeau JH, Aitman TJ (2002) Finding genes that underlie complex traits. Science **298:** 2345-2349

Hilhorst H, Toorop P (1997) Review on dormancy, germinability, and germination in crop and weed seeds. Advances in Agronomy **61:** 111-165

Hilhorst H.W.M. F-SWE, Buitink J., Bolingue W. & Leubner-Metzger G. (2010) Dormancy in plant seeds. In Dormancy and Resistance in Harsh Environments (eds E. Lubzens, J. Cerda & M. Clark), pp. 43–67. Springer, Berlin, Germany.

Hilhorst HWM (2007) Definition and hypotheses of seed dormancy. In: K.J. Bradford and H. Nonogaki (Eds.) Chapter 4: Seed development, dormancy and germination. Annual Plant Reviews Vol. 27. Blackwell Publishing, Sheffield, UK**:** 50-71

Jansen RC, Nap JP (2001) Genetical genomics: the added value from segregation. TRENDS in Genetics **17:** 388-390

Jurado E, Westoby M (1992) Seedling growth in relation to seed size among species of arid australia. Journal of Ecology **80:** 407-416

Kadarmideen HN, Von Rohr P, Janss LLG (2006) From genetical genomics to systems genetics: potential applications in quantitative genomics and animal breeding. Mammalian genome **17:** 548-564

Kazmi RH, Khan N, Willems LA, AW VANH, Ligterink W, Hilhorst HW (2012) Complex genetics controls natural variation among seed quality phenotypes in a recombinant inbred population of an interspecific cross between Solanum lycopersicum x Solanum pimpinellifolium. Plant, Cell & Environment **35:** 929-951

Kazmi RH, Khan N, Willems LA, AW VANH, Ligterink W, Hilhorst HW (2012) Complex genetics controls natural variation among seed quality phenotypes in a recombinant inbred population of an interspecific cross between Solanum lycopersicum x Solanum pimpinellifolium. Plant, Cell & Environment **35:** 929-951

Keurentjes J, Sulpice R, Gibon Y, Steinhauser M, 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

Keurentjes JJB, Fu J, Terpstra IR, Garcia JM, Van Den Ackerveken G, Snoek LB, Peeters AJM, Vreugdenhil D, Koornneef M, Jansen RC (2007) Regulatory network construction in Arabidopsis by using genome-wide gene expression quantitative trait loci. Proceedings of the National Academy of Sciences **104:** 1708-1713

Khan N, Kazmi RH, Willems LAJ, van Heusden AW, Ligterink W, Hilhorst HWM (2012) Exploring the Natural Variation for Seedling Traits and Their Link with Seed Dimensions in Tomato. PLoS One **7:** e43991

Koornneef M, Bentsink L, Hilhorst H (2002) Seed dormancy and germination. Current Opinion in Plant Biology **5:** 33-36

Lee W, Fenner M (1989) Mineral nutrient allocation in seeds and shoots of twelve Chionochloa species in relation to soil fertility. The Journal of Ecology **77:** 704-716

Li Y, Álvarez OA, Gutteling EW, Tijsterman M, Fu J, Riksen JAG, Hazendonk E, Prins P, Plasterk RHA, Jansen RC (2006) Mapping determinants of gene expression plasticity by genetical genomics in C. elegans. PLoS genetics **2:** e222

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

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

Loudet O, Chaillou S, Camilleri C, Bouchez D, Daniel-Vedele F (2002) Bay-0× Shahdara recombinant inbred line population: a powerful tool for the genetic dissection of complex traits in Arabidopsis. TAG Theoretical and Applied Genetics **104:** 1173-1184

Loudet O, Gaudon V, Trubuil A, Daniel-Vedele F (2005) Quantitative trait loci controlling root growth and architecture in Arabidopsis thaliana confirmed by heterogeneous inbred family. TAG Theoretical and Applied Genetics **110:** 742-753

Mackay TFC, Stone EA, Ayroles JF (2009) The genetics of quantitative traits: challenges and prospects. Nature Reviews Genetics **10:** 565-577

Malamy JE, Benfey PN (1997) Down and out in Arabidopsis: The formation of lateral roots. Trends in Plant Science **2:** 390-396

Mexal JG, Landis TD (1990) Target seedling concepts: height and diameter. *In* Proceedings, Target Seedling Symposium, Combined Meeting of the Western Forest Nursery Associations, pp 15-17

Meyer RC, Steinfath M, Lisec J, Becher M, Witucka-Wall H, Torjek O, Fiehn O, Eckardt A, Willmitzer L, Selbig J, Altmann T (2007) The metabolic signature related to high plant growth rate in Arabidopsis thaliana. Proceedings of the National Academy of Sciences of the United States of America **104:** 4759-4764

Morley M, Molony CM, Weber TM, Devlin JL, Ewens KG, Spielman RS, Cheung VG (2004) Genetic analysis of genome-wide variation in human gene expression. Nature **430:** 743-747

Nieuwhof M, Garretsen F, Oeveren J (1989) Maternal and genetic effects on seed weight of tomato, and effects of seed weight on growth of genotypes of tomato (Lycopersicon esculentum Mill.). Plant Breeding **102:** 248-254

Ouyang XR, van Voorthuysen T, Toorop PE, Hilhorst HWM (2002) Seed vigor, aging, and osmopriming affect anion and sugar leakage during imbibition of maize (Zea mays L.) caryopses. International Journal of Plant Sciences **163:** 107-112

Ranal MA, Santana DG (2006) How and why to measure the germination process? Revista Brasileira de Botânica **29:** 1-11

Ruffel S, Krouk G, Coruzzi GM (2010) A systems view of responses to nutritional cues in Arabidopsis: toward a paradigm shift for predictive network modeling. Plant Physiology **152:** 445-452

Shendure J, Ji H (2008) Next-generation DNA sequencing. Nature Biotechnology **26:** 1135-1145

Spano C, Buselli R, Castiglione MR, Bottega S, Grilli I (2007) RNases and nucleases in embryos and endosperms from naturally aged wheat seeds stored in different conditions. Journal of Plant Physiology **164:** 487-495

Steuer R (2006) Review: on the analysis and interpretation of correlations in metabolomic data. Briefings in bioinformatics **7:** 151-158

Sulpice R, Trenkamp S, Steinfath M, Usadel B, Gibon Y, Witucka-Wall H, Pyl ET, Tschoep H, Steinhauser MC, Guenther M (2010) Network analysis of enzyme activities and metabolite levels and their relationship to biomass in a large panel of Arabidopsis accessions. The Plant Cell Online **22:** 2872-2893

Tanksley SD (1993) Mapping polygenes. Annual Review of Genetics **27:** 205-233

Toubiana D, Semel Y, Tohge T, Beleggia R, Cattivelli L, Rosental L, Nikoloski Z, Zamir D, Fernie AR, 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

Tuinstra M, Ejeta G, Goldsbrough P (1997) Heterogeneous inbred family (HIF) analysis: a method for developing near-isogenic lines that differ at quantitative trait loci. TAG Theoretical and Applied Genetics **95:** 1005-1011

Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB (2008) Alternative isoform regulation in human tissue transcriptomes. Nature **456:** 470-476

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics **10:** 57-63

Yin JQ, Zhao RC, Morris KV (2008) Profiling microRNA expression with microarrays. Trends in biotechnology **26:** 70-76

Zhang HM, Jennings A, Barlow PW, Forde BG (1999) Dual pathways for regulation of root branching by nitrate. PNAS**96:** 6529-6534.

Summary

Revolutionary changes in crop agriculture over the last century have transformed the status of seeds of major crops and vegetable varieties from a commodity to an intellectual property. The axiom of farming that desires bountiful harvests often lacks access to high quality seeds. Thus, seeds have become both a delivery system and a valuable asset in our modern, mechanized, commercialized, crop agriculture.

The performance of seeds is determined by three linked and interactive components that constitute a performance triangle of genetics, physiological quality and the environment. So far, there has been little or no discussion about the genetic analysis of seed and seedling traits in tomato at a systems level. To the best of our knowledge, the present study is the first system analysis of the genetics of seed and seedling traits, adding to a growing body of information on tomato seed quality. With the aim of improving the production of high-quality tomato seeds, a multidisciplinary study (physiology, genetics and genomics) was undertaken to develop and evaluate methods for improving the percentage, rate and uniformity of germination and early seedling development, and for increasing the range of environmental conditions for germination. The route of the research was as follows. Primarily, we explored the natural variation present in a *S. lycopersicum* x *S. pimpinellifolium* RIL population to dissect the molecular-genetic mechanisms controlling seed quality. In contrast to lots of plant traits, which are frequently controlled by the action of small-to-large-effect genes that follow classic Mendelian inheritance, our study suggests that seed quality traits, are primarily quantitative and genetically complex. Chapter 2 describes the use of QTL mapping to detect the genomic regions, affecting complex traits in tomato, mainly focusing on traits of direct economic interest in terms of germination potential. This

study has shown that, a diverse set of phenotypic traits maps to the 12 chromosomes on the tomato genome influencing seed quality phenotypes under non-stress, as well as salt-, osmotic-, cold-, high-temperature- and oxidative stress conditions. The QTL approach appeared to be valuable not only in elucidating the genetics, but also the physiological background of the seed quality phenotypes. Both stress-specific and non-stress specific QTLs control the germination process under different environmental conditions in tomato.

The term seed quality has long claimed a very broad meaning that essentially embraces all of the genetic, physiological and physical attributes of seeds. It can be argued that some aspects or attributes of seeds e.g. physical purity, vigor, viability, a high rate of germination, and production of normal seedlings under various environmental (stress) conditions are of great importance in crop production, exceed the boundaries of quality and are best described and understood in terms of performance. Good seedling establishment and seedling vigor are essential for sustainable and profitable crop production and are considered to constitute the most critical stage of a developing crop. Numerous QTLs were identified along the tomato genome in the *S. lycopersicum* x *S. pimpinellifolium* RIL population for seed, seedling and root system architecture traits. These traits showed overlapping QTLs for different seed and seedling traits with generally the same genetic direction of the QTLs supported by the correlation in the phenotypic values of these traits. This indicates a strong relationship among seed- and seedling vigor whereby seed size and weight ultimately affect the initial growth of the main root and the upward growth of the shoot (Chapter 3). However, a relationship between seed weight and seed

germination could not be established in the present study. Interestingly, seed weight had a strong effect on seedling vigor and these results are of great importance for the isolation of the corresponding genes and elucidation of the underlying mechanisms. Another important finding was that germination performance and seed size are controlled by different independent genetic loci. Robust QTL mapping with SNP-based linkage maps offers a much-improved estimation of the genetic architecture of a tomato genome in terms of the magnitude of QTL effects, QTL-environment interactions and putative pleiotropy. Thus, this approach offers a way in which simultaneous improvement of these traits and progress towards identifying the underlying genetic mechanisms may be realized in the near future.

Though previous solutions to issues associated with seed quality phenotypes seemed promising, none have utilized the integration of genomic, phenotypic and metabolic datasets to understand seed quality in tomato. Thus, the integration of metabolic and genomic analysis contributed to a comprehensive biological understanding of observed phenotypic differences between RILs of *S. lycopersicum* x *S. pimpinellifolium.* The metabolite abundance profiling in tomato seeds, as described in Chapter 4, uncovered extensive genetic variation in metabolite abundance in the RIL population and also comprehensively demonstrated the integration of different datasets. Here we described, for the first time, a generalized genetical genomics (GGG) model in tomato seeds that incorporates genetics, as well as environmental effects, and we applied this approach to mapping quantitative trait loci (QTLs) and QTLs that are the result of interaction between the genetics and environmental changes (Genetic x Environmental QTLs). In particular, it uses chosen environmental perturbations (different seed developmental stages, i.e. dry

and 6h imbibed seeds) in combination with the analysis of genetic variation present in the RIL population, to study the change of metabolites over the multiple environments and to identify genotype-by-environment interactions. QTL analysis with 160 detected metabolites in the RIL population resulted in QTL hotspots for over two-thirds of the metabolites.

Novel findings were also observed; particularly the transition from dry to 6h imbibed seeds was associated with programmed metabolic switches, showing various metabolites that were synthesized in accordance with demand and possible utilization. This suggests that metabolic pathways proceeding during the 6h imbibition has to substantiate the demand for certain metabolites, in particular for amino acids, as the majority of them clustered together. The metabolites most relevant to germination traits were extracted using multivariate statistics (canonical correlation). Strong evidence of canonical correlation was found between metabolites and seed quality traits, revealing a close link between seed quality phenotypes and a specific combination of metabolites. The result to emerge from the analysis was that there is no single metabolite associated with germination traits; instead, a group of metabolites appears to explain the total variation for a certain phenotypic trait. The question of metabolic regulatory mechanisms, was further disentangled by constructing a complementary overview of the metabolite network obtained by extraction of all significant trait-trait correlations. The topologies of the identified metabolic correlation modules may contribute to enhancing our understanding of the role of highly connected metabolites in these metabolic networks and their specific functions in seed-developmental biochemical pathways. Densely connected metabolites were extracted using graph clustering from correlation networks, and the clusters were evaluated by biochemical-pathway enrichment analysis. The result from

this study suggests that the number of significant correlations varied among individual metabolites and that the obtained clusters were significantly enriched for metabolites involved in specific biochemical pathways.

Immortal populations are a powerful tool for use in QTL analysis to explore the interaction of organisms with their environment. High-precision mapping leading to fine mapping; responsible for complex trait variation was achieved by the construction of NILs. Advantage was taken of residual heterozygosity left at the F_8 generation to find lines that still segregated only at the region around the QTL of interest, a type of NILs called HIFs (Chapter 5). A number of QTLs were confirmed using HIFs affecting germination phenotypes under non-stress (control) and stress conditions at the bottom of chromosome 6, as well as QTLs for various germination phenotypes on chromosome 8. This confirmation with respect to the variation observed in the previous phenotyping experiments with the RILs was relevant and repeatable. These genomic regions regulating seed germination phenotypes under the various conditions are plausible candidates for further fine mapping, isolation of the casual genes and characterization and molecular dissection of the pathways in which they function.

This thesis gives an account of the integration of genotyping, phenotyping and a molecular phenotype using metabolomics in generating a novel understanding of seed phenotypes and their interaction with the environment. In summary, the integration of phenotypic and metabolomics data has facilitated the identification of potential biomarkers. Presuming further analytical and biological validation these markers may be useful in future seed quality studies, for example by additional integration of transcriptomics data. These follow-up studies may enable further

integrative pathway mapping; in this way particular classes of genes may be identified that contribute to our understanding of the complex nature of seed quality. The integration of multiple layers of biological information from DNA to phenotype to environment will create system-level models of seed quality.

Samenvatting

Gedurende de afgelopen eeuw hebben revolutionaire veranderingen in de landbouw de status van zaden van de voornaamste landbouw gewassen veranderd van enkel een handelsartikel naar een intellectueel eigendom. Vaak zijn er geen kwalitatief hoogwaardige zaden beschikbaar om te voldoen aan de steeds groter wordende vraag naar hogere opbrengst en betere kwaliteit van land- en tuinbouwproducten. Daarom zijn kwalitatief hoogwaardige zaden een waardevolle aanwinst geworden voor onze moderne, mechanische en gecommercialiseerde landbouw.

De zaadkwaliteit wordt bepaald door drie factoren, te weten genetica, fysiologie en omgevingsfactoren. Tot dusverre is er weinig systematisch onderzoek gedaan naar de genetische factoren die een invloed hebben op zaad- en kiemplanteigenschappen in tomaat. Voor zover wij kunnen nagaan is de in dit proefschrift beschreven studie de eerste systematische analyse van de genetica van eigenschappen van tomatenzaad en kiemplant en draagt daarmee bij aan de groeiende hoeveelheid informatie over zaadkwaliteit.

We hebben geprobeerd om met een multidisciplinaire aanpak methoden te ontwikkelen en te valideren voor de verbetering van het kiemingspercentage, de kiemsnelheid en uniformiteit, evenals vroege kiemplantontwikkeling onder invloed van verschillende omgevingsfactoren. Voor dit onderzoek hebben we de volgende stappen ondernomen: in eerste instantie hebben we de genetische variatie onderzocht die te vinden is in een genetische karteringspopulatie (recombinante inteelt lijnen; RIL) van *Solanum lycopersicum* en *Solanum pimpinellifolium* om de moleculaire mechanismen die zaadkwaliteit bepalen te ontrafelen.

In tegenstelling tot vele planteneigenschappen die vaak worden gecontroleerd door de werking van genen die een klassieke Mendeliaanse

overerving laten zien, suggereren onze resultaten dat kwaliteitskenmerken van zaad voornamelijk kwantitatief van aard en genetisch complex zijn. Hoofdstuk 2 beschrijft het gebruik van QTL (quantitative trait loci) analyse om de genomische gebieden te vinden die invloed hebben op complexe eigenschappen in tomaat met de nadruk op die eigenschappen die een direct economische belang hebben met betrekking tot kiemingspotentiaal. Deze studie heeft ervoor gezorgd dat we een goed overzicht hebben van de QTLs die invloed hebben op eigenschappen van zaadkwaliteit onder normale omstandigheden en onder zoute, osmotische, koude, warme en oxidatieve stress omstandigheden. Met behulp van deze QTL analyse hebben we niet alleen de genetische regulatie van deze eigenschappen kunnen ophelderen, maar ook de fysiologische achtergrond van deze kenmerken van zaadkwaliteit.

Goede ontwikkeling en groeikracht van kiemplanten zijn van essentieel belang voor duurzame en winstgevende landbouw en worden in het algemeen gezien als de meest kwetsbare fase in de ontwikkeling van planten. In Hoofdstuk 3 beschrijven we de identificatie van verschillende QTLs voor eigenschappen van zaad-, kiemplant- en wortelarchitectuur. Voor verschillende zaad- en kiemplantkenmerken vonden we hierbij overlappende QTLs, die meestal een gelijke genetische richting hadden en ook de fenotypische waarden voor deze eigenschappen lieten sterke correlaties zien. Dit duidt op een sterke relatie tussen zaad- en kiemplanteigenschappen, waarbij zaadgrootte en gewicht een effect hebben op de initiële groei van de wortel en de stengel van kiemplanten. Deze informatie zal de isolatie van de corresponderende genen vergemakkelijken en dit zal helpen om de onderliggende moleculaire mechanismen te verklaren. In tegenstelling tot deze correlatie hebben we in onze studie geen verband gevonden tussen zaadgrootte en

kiemingseigenschappen. In verband hiermee hebben we ook kunnen vaststellen dat zaadgrootte en kiemingseigenschappen worden gereguleerd door onafhankelijke genoomregio's. Robuuste QTL analyse met behulp van een op SNPs (single nucleotide polymorphism) gebaseerde genetische kaart heeft er voor gezorgd dat we een goede inschatting konden maken van de genetische architectuur van het tomatengenoom met betrekking tot de grootte van de QTL effecten, de interacties tussen de QTLs en de omgeving en mogelijke pleiotropie. Deze aanpak maakt het mogelijk dat de hier bestudeerde eigenschappen gelijktijdig kunnen worden verbeterd door middel van plantenveredeling en zal helpen in de identificatie van de genetische mechanismen die deze eigenschappen controleren.

Alhoewel eerdere oplossingen om problemen in verband met zaadkwaliteit op te lossen veelbelovend lijken, hebben ze geen van allen de integratie van genomische, fenotypische en metabolische data gebruikt om zaadkwaliteit in tomaat te bestuderen. De integratie van genomische en metabolische analyses heeft er aan bijgedragen dat we nu een completer inzicht hebben in de biologie achter de door ons gemeten fenotypische verschillen tussen de *S. lycopersicum* x *S. pimpinellifolium* RILs. Het meten van metabolietniveaus in tomaten zaden zoals beschreven in Hoofdstuk 4 heeft laten zien dat er uitgebreide variatie bestaat voor de niveaus van de metabolieten in de bestudeerde RILs. Het meten van de metabolietniveaus in de RIL lijnen is voor het eerst in tomaat gebeurd met een gegeneraliseerde experimentele opzet. Deze opzet maakt het mogelijk om een 'genetisch omics' experiment uit te voeren en wel gelijktijdig voor verschillende ontwikkelingsstadia. Naast informatie over de genetische controle van metaboliet niveaus geeft een dergelijke opzet ook additionele informatie over effecten van ontwikkelingsstadia op metabolietniveau en op de interactie tussen genetische en ontwikkelings-effecten. In dit onderzoek

hebben we gekeken naar de metabolietniveaus in droge en 6 uur geimbibeerde zaden. QTL analyse van de 160 gedetecteerde metabolieten resulteerde in de ontdekking van QTLs voor tweederde van de metabolieten. Verdere analyse resulteerde in de ontdekking dat de transitie van droog naar 6 uur geimbibeerd zaad is geassocieerd met metabolische omschakelingen waarbij verschillende metabolieten die nodig zijn voor de gedurende imbibitie opgestarte processen worden aangemaakt. Vooral de hoeveelheid van meerdere aminozuren ging omhoog en deze gaven ook een sterke clustering te zien bij analyse van de data. Met behulp van canonische correlatie analyse hebben we de metabolieten geïdentificeerd die het meeste relevant leken voor de geobserveerde kiemeigenschappen. De canonische correlatie analyse onthulde een nauw verband tussen combinaties van metabolieten met verschillende zaadkwaliteits-eigenschappen. Er werden geen metabolieten gevonden die op zich zelf associeerden met kiemeigenschappen, maar groepen van metabolieten konden wel een goede verklaring geven voor de gevonden variatie in bepaalde kiemeigenschappen. Metabolische netwerken werden verder geanalyseerd door de extractie van alle significante eigenschap-eigenschap correlaties. De topologie van de op deze wijze verkregen modules van metabolische correlatie kunnen bijdragen aan het begrijpen van de specifieke functies van de metabolieten in metabolische routes van de zaadontwikkeling.. Sterk verbonden metabolieten werden met behulp van grafiekclustering van correlatienetwerken geëxtraheerd en de zo gevonden clusters werden geanalyseerd met behulp van biochemische-route verrijking analyse. Het resultaat van dit onderzoek suggereert dat het aantal significante correlaties varieerde tussen individuele metabolieten en dat de verkregen clusters aanzienlijk werden verrijkt voor metabolieten betrokken bij specifieke biochemische routes.

Onsterfelijke populaties, zoals RIL populaties, zijn een krachtig hulpmiddel om de interactie van organismen met hun omgeving te bestuderen. Aangezien de genomische gebieden die met behulp van QTL analyses worden geïdentificeerd nog behoorlijk groot zijn, is verder kartering nodig om deze gebieden te verkleinen en uiteindelijk één causaal gen te kunnen aanwijzen. De methode om dit te doen is door gebruik te maken van zogenaamde NILs (near isogenic lines). Wij hebben gebruik gemaakt van resterende heterozygotie in onze RILs om zogenaamde HIFs (heterogenous inbred families) te maken. Dit betreft een specifiek soort NILs. In Hoofdstuk 5 bespreken we de isolatie van deze HIFs en de bevestiging van de QTLs, met behulp van deze lijnen, die we voor verschillende eigenschappen hebben gevonden. Onderaan chromosoom 6 hebben we op deze manier een QTL bevestigd voor verschillende kiemeigenschappen onder controle- en stress-omstandigheden. Verder hebben we op chromosoom 8 een QTL voor kieming onder oxidatieve stress bevestigd. Deze resultaten laten zien dat de aanpak voor het vinden van de QTLs relevant en betrouwbaar is geweest. De hier beschreven HIFs zijn uitstekend geschikt als uitgangsmateriaal voor meer nauwkeurigere kartering van de regio's met als uiteindelijke doel de isolatie van de causale genen en de verdere karakterisatie en moleculaire ontrafeling van de routes waarin ze functioneren.

Dit proefschrift geeft een overzicht van de integratie van genotypering, fenotypering en moleculair fenotypes met behulp van metabolietanalyse om een beter begrip van zaadfenotypes en hun interactie met de omgeving te genereren. Samenvattend heeft de integratie van fenotypische en metabolische data, potentiële biomerkers geïdentificeerd. Na verdere analytische en biologische validatie van deze merkers kunnen ze bruikbaar zijn voor toekomstige studies naar

zaadkwaliteit. Vervolgonderzoek, zoals de additionele integratie van genexpressie data met de in dit proefschrift beschreven data, zullen additionele routes voor de controle van zaadkwaliteit aan het licht brengen. Op deze manier kunnen specifieke groepen van genen worden geïdentificeerd die zullen bijdragen aan ons begrip van de complexiteit van zaadkwaliteit. De integratie van meerdere lagen van biologische informatie van DNA naar fenotype naar omgeving zal uiteindelijk de formulering van modellen voor zaadkwaliteit op systeemniveau mogelijk maken.

Acknowledgements

I would like to express the deepest appreciation to Henk Hilhorst, who has the attitude and the substance of a genius: he continually and convincingly conveyed a spirit of adventure in regard to research and brought unique perspectives. I appreciate all his contributions of time, ideas, and supervision to make my Ph.D. experience productive and stimulating.

I would like to thank my supervisor, Wilco Ligterink. I am very appreciative of his generosity with his time, advice, and references, to name a few of his contributions. The joy and enthusiasm he has was contagious and motivational for me, even during tough times in the Ph.D. pursuit. Without his guidance and persistent help this dissertation would not have been possible. I consecrate my sincere thanks to my promoter Professor Harro Bouwmeester for his valuable suggestions and guidance. I gratefully acknowledge Maarteen Koorneef, Joost Keurentjes, Bas Dekkers and Leonie Bentsink, who always give me suggestions and advise me when I was stuck.

The members of the Plant Physiology and in particular Wageningen Seed Lab have contributed immensely to my personal and professional time at Wageningen. The group has been a source of friendships as well as good advice and collaboration. I would like to acknowledge Noorullah and Ronny as we work close together during all these years, and I very much appreciated their enthusiasm, intensity, willingness to help each other in time of need. Other past and present group members that I have had the pleasure to work with or alongside the numerous guest and rotation students who have come through the lab.

I give my heartfelt thanks to my friends, Ishtiaq Khaliq, Awais Khan, Liaqat Ali, Abrar Nisar, Imran Ali, Nisar Akbar Khan, Nadeem Mughal, Sher Afgan, for their encouragement and moral support. I am appreciative of to all the people in particular, I would like to thank these individuals: Imran Haider, Zeshan Hassan, Mohamed El Soda, Cihan Tas Emma, Geke, Leonardo, Lisanne, Margriet, Marta, Matthijs, Saskia, Sara, Veronika, Ingrid, Ati for making my time at Wageningen a more enjoyable experience.

In addition, a thank you to all the friends and groups at Stichting OtherWise, who introduced me to what it seemed like an eternity hunting toil of activities of my interest. Otherwise offered me a way to explore an interest in a dynamic social and political environment that I had really always harbored. It ended up being the best experience of my life as it gave me opportunity to perceive Dutch as well people across the globe.

I would like to extend my deepest gratitude to my family: my parents Abjee and Maanjee, my brothers Wajid Kazmi, Amjad Kazmi, Arshad Kazmi, Hamid Kazmi and Shahid Kazmi, my sisters Fakhra Tabassum, Yasira Tabassum and Afia Hassan, and my nieces and nephew Iraj Fatima, Abiha Fatima and Ibtissam Ali. They always have and provided unwavering love and encouragement.

Thank you for believing in me.

Finally, I would like to express my deepest gratitude to Leo Willems for everything from technical support to hilarious jokes. Thank you for being there for me from the very beginning.

And so let us always meet each other with a smile, for the smile is the beginning of love, and once we begin to love each other naturally we want to do something (Mother Teresa).

Be good, be kind, and be thankful.

Rashid Kazmi

Biography - Curriculum vitae

Rashid was born and raised in oath of Himalayas in a relatively small town Mujhoi about 10 miles from Muzaffarabad, Jammu Kashmir (Pakistan); famous for its forests, rivers surrounded by majestic mountains. While growing up, my family and friends helped me develop a wide range of interests but especially sports, outdoor activities, wildlife, and travel.

EDUCATION & QUALIFICATIONS

Sep 2008- Dec 2012 **Wageningen Seed Lab, Wageningen UR, The Netherlands**
PhD in Plant Physiology (Seed Biology)
Thesis Title: Genes for seed quality:
"Integrating physiology and genetical genomics to mine for seed quality genes in tomato"
Supervisors: Dr. Henk W.M. Hilhorst, Dr. Wilco Ligterink
External-Supervisor: Prof. Dr Maarten Koornneef

Sep 2004 – Feb 2006 **University of Sussex, Brighton, United Kingdom**
MSc in Bioinformatics

Jan 2003 –Sep 2004 **London City Institute, United Kingdom**
Diploma in Computer Science

Dec 1997 – Apr 2002 **University of Azad Jammu & Kashmir, Pakistan**
BSc (Hons)- Plant Breeding & Genetics (1st Class)

A-levels: Physics, Chemistry , Biology, English
GCSEs: 8 at grades A to C, including Maths and English

RELATED WORK EXPERIENCE

Sep 2006 – July 2007 **University of Sussex, United Kingdom**
Research Assistant
Voluntary assignment in the Department of Biology & Environment Sciences

Sep 2000 – Oct 2002 **University of Azad Jammu & Kashmir, Pakistan**
Research Officer

OTHER WORK EXPERIENCE

Sep 2010 – Present Stichting OtherWise, Wageningen, The Netherlands
Chairman
Voluntary position

Feb 2003 – Nov 2003 Select Service Partner, Gatwick Airport, UK
Customer Sales Assistant,

Nov 1999 – Jun 2000 Sukhi Development Foundation, Pakistan
Biodiversity Management Adviser

SPECIAL ACHIEVEMENT
Certificate of Recognition for an outstanding essay presented by Secretary Ministry of Agriculture, Government of Pakistan and representative of Food and Agriculture Organisation of the United Nations, October, 1999

Related to this thesis

- **Kazmi R. H.**, Khan N., Willems L.A.J., van Heusden A.W., Ligterink W., Hilhorst H.W.M. (2011) 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(5), pp. 929-951

- Khan N, **Kazmi R. H.**, Willems LAJ, van Heusden AW, Ligterink W, et al. (2012) Exploring the Natural Variation for Seedling Traits and Their Link with Seed Dimensions in Tomato. PLoS ONE 7(8): e43991. doi:10.1371/journal.pone.0043991

- **Kazmi R. H.**, Khan N., Joosen R.V.L, Willems L.A.J., Ligterink W., Hilhorst H.W.M. (2013) Metabolomics: The link between genotype, phenotype and metabolome in tomato seed quality attributes (in preparation).

- **Kazmi R. H.**, Khan N.,Willems L.A.J., Ligterink W., Hilhorst H.W.M. (2013). Dissection of the complex phenotypes of seed quality on tomato chromosomes 6 and 8: HIFs (heterogeneous inbred families) confirmation (in preparation).

Other publications

- **Kazmi R. H.,** Khan M. Q. and Abbasi M. K. (2003) Yield and Yield Components of Wheat subjected to water stress under rainfed conditions in Pakistan, Acta Agronomica Hungrica, 5 (3), pp.315-323

- Abbasi M. K., **Kazmi R. H**. and Khan M. Q. (2003) Growth Performance and Stability analysis of some wheat (*Triticum aestivum* L.) genotypes subjected to water stress at Rawalakot, Azad Jammu and Kashmir, Archives of Agronomy and Soil Science, Vol. 49, pp.415-426

Education Statement of the Graduate School Experimental Plant Sciences

Issued to: **Rashid Hussain Kazmi**

Date: **20-March-2013**

Group: **Laboratory of Plant Physiology, Wageningen UR**

1) Start-up phase	*date*	*cp*
▶ **First presentation of your project** Genes for seed quality: Integrated physiology and genetical genomics to mine for seed quality genes in tomato	Jan 12, 2009	1.5
▶ **Writing or rewriting a project proposal** Genes for seed quality: Integrated physiology and genetical genomics to mine for seed quality genes in tomato	Dec 15, 2008	6.0
▶ **Writing a review or book chapter**		
▶ **MSc courses**		
▶ **Laboratory use of isotopes**		
Subtotal Start-up Phase		7.5

2) Scientific Exposure	*date*	*cp*
▶ **EPS PhD Student Days**		
EPS PhD student day, Museum Naturalis, Leiden	Feb 26, 2009	0.3
EPS PhD student day, Uithof, Utrecht	Jun 1, 2010	0.3
EPS PhD student day, Wageingen	May, 2011	0.3
▶ **EPS Theme Symposia**		
EPS theme 4 'Genome Plasticity', Wageningen University, The Netherlands	Dec 12, 2008	0.3
EPS theme 2 'Interactions between Plants and Biotic Agents', Utrecht University, The Netherlands	Jan 22, 2009	0.3
EPS theme 1 'Developmental Biology of Plants', Leiden University, The Netherlands	Jan 30, 2009	0.3
EPS theme 4 'Genome Plasticity', Radboud University, The Netherlands	Dec 1, 2009	0.3

EPS theme 1 'Developmental Biology of Plants', Wageningen University, The Netherlands	Jan 28, 2010	0.3
EPS theme 3 'Metabolism and Adaptation', Wageningen University, The Netherlands	Feb, 10, 2011	0.3
EPS theme 1 'Developmental Biology of Plants', Wageningen University, The Netherlands	Jan 19, 2012	0.3
EPS Theme 3 symposium, 'Metabolism and Adaptation', Utrecht	Apr, 26, 2012	0.3
EPS Theme 3 symposium, 'Metabolism and Adaptation', Lieden	Feb, 19, 2010	0.3
NWO Lunteren days and other National Platforms		
ALW meeting 'Experimental Plant Sciences', Lunteren	Apr 6-7, 2009	0.6
ALW meeting 'Experimental Plant Sciences', Lunteren	Apr 19-20, 2010	0.6
ALW meeting 'Experimental Plant Sciences', Lunteren	Apr 4-5, 2011	0.6
ALW meeting 'Experimental Plant Sciences', Lunteren	Apr 2-3, 2012	0.6
Seminars (series), workshops and symposia		
Seminar: Prof. Dr. Zhenbi Yang	Jun, 23 2008	0.1
Seminar: Prof. Dr. Hong-Qing Ling	Jul, 21 2008	0.1
Seminar: Prof. Jian-Kang Zhu	Nov, 3 2008	0.1
Symposium: The Schilperoort Lectures: Success Stories of Entrepreneurial Scientists, Wageningen University, The Netherlands	Nov, 5 2008	0.3
Seminar: Science From an Editor's View, by Dr. Pamela J. Hines, Wageningen University, The Netherlands	Nov, 6 2008	0.1
Seminar: Prof. Dr. Sjef Smeekens	Nov, 27 2008	0.1
Symposium: New Opportunities for Conservation Genetics with Genome Wide Information, Wageningen, The Netherlands	Dec, 8 2008	0.3
Mini Symposium: Bibliometrics at Wageningen UR, Wageningen University, The Netherlands	Nov , 27 2008	0.1
Seminar: Dr. Hiro Nonogaki	Sep, 17 2009	0.1
EPS Symposium: Ecology and Experimental Plant Sciences, Wageningen University, The Netherlands	Sep, 22 2009	0.3
Farewell Symposium: Dr. Pim Zabel ''Art Meets Science'', Wageningen University, The Netherlands	Oct, 16 2009	0.3
Symposium: National EcoGenomics Day, Amsterdam, The Netherlands	Apr, 21 2010	0.3
Seminar: Dr. John Yoder	Aug, 26 2010	0.1
Seminar: Dr. David Baulcombe	Sep, 27 2010	0.1
Seminar: SNIP Detection by Allumina Sequencing, by Dr. Robert Kraus	Oct, 7 2010	0.1
Wageningen UR Sequencing Seminar, Wageningen University, The Netherlands	Dec, 7 2011	0.3
Seminar: Prof. Dr. Steffen Abel	Mar, 20 2011	0.1
Seminar: Prof Graham Seymour	Jan, 24 2012	0.1

Seminar : Prof Steven Penfield	Jun, 12 2012	0.1
Seminar : Prof Jill M. Farant	Jun, 26 2012	0.1
Seminar : Dr. Lauren McIntyre	Sep, 17 2012	0.1
Seminar : Prof Ruth Finkelstein	Nov, 14 2012	0.3
Seminar: Dr. Aaron Fait	Dec, 4 2012	0.3
Mini-Symposium "Plant Molecular Physiology", Utrecht	Nov, 20 2012	0.3
Seminar series, Plant Physiology-Genetics, Wageningen University, The Netherlands	2008-2012	3.0
Seminar plus		
International symposia and congresses		
3rd Workshop on Molecular Aspects of Seed Dormancy and Germination, York, UK	Jul, 18-21 2010	1.5
10th ISSS Workshop on Seed Biology, Salvador (Bahia State), Brazil	Apr, 13-18 2011	1.2
ASPB "Plant Biology 2012" meeting, Austin, USA	Jul, 20-24 2012	1.5
3rd Joint PhD Retreat, Orsay, France	Jul, 5-8 2011	0.9
2nd PhD retreat, Cologne, Germany	Apr, 15-17 2010	0.9
Presentations		
Poster: EPS PhD Day, Wageningen	May, 20 2011	0.7
Poster: 2nd PhD retreat, Cologne, Germany	Apr, 15-17 2010	0.7
Poster: ALW meeting 'Experimental Plant Sciences', Lunteren	Apr, 6-7 2009	0.7
Poster: ALW meeting 'Experimental Plant Sciences', Lunteren	Apr, 19-20 2010	0.7
Poster: ALW meeting 'Experimental Plant Sciences', Lunteren	Apr, 2-3 2012	0.7
Poster: ASPB "Plant Biology 2012" meeting, Austin, USA	Jul, 20-24 2012	0.7
Oral Presentation: ALW meeting 'Experimental Plant Sciences', Lunteren	Apr, 5 2011	0.7
Oral Presentation: 10th ISSS Workshop on Seed Biology, Salvador, Brazil	Apr, 13-18 2011	0.7
Oral Presentation: Genes for seed quality: 2nd PhD retreat, Paris, France		0.7
IAB interview	Feb, 18 2011	0.7
Excursions		
Subtotal Scientific Exposure		25.5

3) In-Depth Studies	*date*	*cp*
EPS courses or other PhD courses		
System Biology: Statistical Analysis of ~Omics Data	Dec, 8-11 2008	1.4
Molecular Phylogenies: Reconstruction & Interpretation	Oct, 19-23 2009	1.5
Master Class Seed Technology	Oct, 26-29, 2009	1.2
Mixed model based QTL mapping in GeneStat	May, 14-16 2012	0.9

Journal club		
Literature discussion: Plant Physiology group, Wageningen University, The Netherlands	2009-2012	3.0
Individual research training		
Subtotal In-Depth Studies		8.0
4) Personal development	*date*	*cp*
Skill training courses		
PhD Competence Assessment	Feb, 24 2009	0.3
Interdisciplinary and Transdisciplinary Research: Intervision and Communication Skills	Jun, 11-15 2009	1.1
Course: Project- and Time Management	Nov, 6,17 Dec. 17, 2009	1.5
EPS Career Event: ExPectationS Day, Wageningen, The Netherlands	Nov, 19 2010	0.3
Workshop: Digital Art, Austin, USA	Jul, 21 2012	0.1
Workshop: Gbrowser, Austin, USA	Jul, 22 2012	0.1
Minisymposia: Education and Outreach, Austin, USA	Jul, 23 2012	0.1
Organisation of PhD students day, course or conference		
Membership of Board, Committee or PhD council		
Subtotal Personal Development		3.5
TOTAL NUMBER OF CREDIT POINTS		**46.0**

Herewith Graduate School declares that the PhD candidate has complied with the educational requirement set by the Education Committee of EPS which comprises a minimum total of 30 credits.

**A credit represents a normative study load of 28 hours of study.*

.....

*Cover page design was conceived from tomato genome project (Nature 485,635–641) and designed by author. http://www.123rf.com/photo_8920311_vector-illustration-of-tomato.html (tomato pictures)

Printed by : Proefschriftmaken.nl Weerdskampweg 15 5222 BA 's-Hertogenbosch, Netherlands
T: +31 (0)73 513 07 04