Genetic and Physiological Quality of Tomato Seed and Seedlings

Noorullah Khan

Thesis committee

Promotor

Prof. Dr. H.J. Bouwmeester
Professor of Plant Physiology

Co-promotors

Dr. H.W.M. Hilhorst
Associate Professor, Laboratory of Plant Physiology
Wageningen University

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

Other members

Prof. Dr. M.E. Schranz, Wageningen University
Prof. Dr. J.J.B. Keurentjes, University of Amsterdam / Wageningen University
Dr. C.H. de Vos, Plant Research International, Wageningen
Dr. P. Spoelstra, Incotec Holding B.V., Enkhuizen

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

Genetic and Physiological Quality of Tomato Seed and Seedlings

Noorullah Khan

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 Tuesday 3 September 2013

at 1.30 p.m. in the Aula.

Noorullah Khan

Genetic and Physiological Quality of Tomato Seed and Seedlings

249 pages

PhD thesis, Wageningen University, Wageningen, NL (2013)

With references, with summaries in English and Dutch

ISBN 978-94-6173-647-5

CONTENTS

Chapter 1

General Introduction

Seed Quality

Seed quality is one of the most important factors to affect the success of a crop (Finch-Savage, 1995) and is thought to be associated with many interlinked physiological and genetic traits (Hilhorst and Koornneef, 2007; Hilhorst et al., 2010). 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 a complex trait and is defined as "the viability and vigour attribute of a seed that enables the emergence and establishment of normal seedlings under a wide range of environments" (Khan et al., 2012). The practical definition of seed quality is determined by the end user and will, therefore, differ substantially, depending on the use of seeds as propagule or commodity. For a farmer or plant grower high quality seeds are those seeds that germinate to a high percentage and establish vigourous seedlings under a wide range of field conditions. On the other hand, high quality seeds for use in the food industry may be seeds with a high starch or oil content or oil seeds with a specific protein or fatty acid composition (Nesi et al., 2008).

Seed quality (for propagation) is determined by a number of physiological processes related to important plant developmental events, 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 comprises many different attributes, including germination characteristics, dormancy, seed and seedling vigour, uniformity in seed size, normal embryo- and seedling morphology, storability, absence of mechanical damage, as well as the ability to develop into a normal and vigourous 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). Because of its complex nature, testing of seed quality is in many cases, at best, an 'educated guess' in order to predict subsequent behavior in the field (Powell and Basra, 2006). Therefore, seed producers have redefined the term 'seed quality' to include important attributes such as 'usable plants' and 'seedling and crop establishment'. The attribute 'usable plants' is one of the major characteristics of seed quality used by seed producers and plant breeders (Ligterink et al., 2012).

Seed quality is mainly acquired during seed development and maturation, and is drastically affected by interactions between the genome and the prevailing environmental conditions. This process is part of the normal adaptation of plants to a varying environment

and is aimed at maximizing the possibility of successful offspring (Huang et al., 2010). As the ultimate performance of a seed is a function of the complex interaction between the genome and the environment, seed quality can be enhanced at all the different steps of the production process. Since it is difficult to influence the production environment, even under greenhouse conditions, plant breeders and seed companies try to acquire the best possible quality of seeds mainly by varying the time and method of harvest, and particularly by post-harvest treatments such as cleaning, sorting, coating and priming and by controlling the storage conditions. However, the genetic component of the interaction between the genome and the environment can be investigated and this variation in genetic adaptation may provide opportunities for plant breeders and seed companies to breed for better seed quality. Despite these opportunities, the genetic regulation of seed quality has hardly been investigated to be used in breeding programs. Although, a few studies have documented some quantitative trait loci (QTLs) associated with germination, storability and stress tolerance in Arabidopsis and tomato (Foolad et al., 2003; Clerkx et al., 2004), a systematic study of the genetics of seed quality is lacking. The present study seeks to discover integrative approaches that can facilitate the understanding of the underlying causes of the complex trait of seed quality. Our objective is to provide new methods for dissecting the genetic components of seed quality by integrating the physiology, genetics, genomics and metabolomics of seeds to identify loci, and subsequently genes, controlling seed quality traits in tomato.

Important Seed Quality Attributes

Seed size variation and its influence on seedling establishment

Among others seed size and mass are important traits determining seed quality(Panthee et al., 2005), which in turn are the most variable traits in the plant kingdom (TeKrony and Egli, 1991; Orsi and Tanksley, 2009). Seed size is a key determinant of evolutionary fitness in plants and is a trait that often undergoes tremendous changes during crop domestication. Seed size is most often quantitatively inherited and seeds range in weight from less than 1 microgram in the Coral-root orchid (*Corallorhiza maculate*) to more than 10 kg in the Coco-de-mer palm (*Lodoicea maldivica*). This large variation in seed size can be observed not only among taxa, but also within taxa. For example, the genus Solanum contains a set of 9 cross-compatible species, closely related to tomato. Despite their close taxonomic affinities, these species show a 10-fold range in seed size, suggesting a rapid rate of evolutionary change. There is typically at least a 10^5 fold variation of seed mass between species within a single area (Westoby et al., 1992; Orsi and Tanksley, 2009). In addition to the variation in seed size among different species, many studies have emphasized that seed size varies significantly within the same species (Michaels et al., 1988) and between different populations and different mother plants and even between different seeds of the same

mother plant. Nevertheless, this variation within species is very small compared to the range across species (Westoby et al., 1996). Many studies have interpreted seed size differences between species by reference to larger seed size being more adaptive under a variety of environmental hazards. However, experimental confirmation of the benefits of large seed size in relation to particular hazards is rare. More experiments are now being reported but a consistent picture has yet to emerge. The reason for this large variation in seed size is not clear. However, evolutionists and ecologists have long observed this great variation and suggested its importance in adaption to different environments (Metz, 1999). With respect to survival there are both risks and benefits for a species to have large or small seed size. Seed size is thought to have evolved as a compromise between producing numerous smaller seeds, each with small resources, and fewer larger seeds, each with more resources. Because seed size trades off with seed number due to limited availability of maternal resources, small seeded species clearly have the advantage in fecundity, but the countervailing advantage of large seeds appears to be their tolerance to stresses such as shade or drought that are present in some but not all regeneration sites (Smith and Fretwell, 1974; Westoby et al., 1992; Metz, 1999; Orsi and Tanksley, 2009; Muller-Landau, 2010). Most of the domesticated crops (e.g. soybean, wheat, maize, sunflower) produce seeds significantly larger than their wild ancestors. It is assumed that crop domestication resulted in increased seed size due to selection for larger seeds in an effort to increase yield and improve harvest efficiency (Broich and Palmer, 1980; Burke et al., 2002; Fuller, 2007; Isemura et al., 2007). However, seed size also increased during domestication in crops other than the ones used for their edible seeds. One example of such crop is the domesticated tomato which produces seeds up to several fold larger than its wild ancestors (Doganlar et al., 2000). The reason for an increase in seed size during domestication of these crops that are not consumed for their seed is unknown. However, it is supposed that an increase in seed size in these species occurred due to indirect selection for greater seedling vigour and germination uniformity under field conditions (Harlan et al., 1973). In tomato, the increase in seed size may be also due to indirect selection for fruit size as seed size is positive correlated with fruit size in tomato (Goldman et al., 1995; Grandillo and Tanksley, 1996).

Despite the evolutionary and agronomic significance of seed weight, relatively little is known about the genetic and molecular mechanisms underlying natural variation in seed size (Doganlar et al., 2000). Most of our knowledge about seed size is confined to quantitative trait mapping studies which have documented a large number of QTLs affecting seed size in a number of non-crop and crop species, including Arabidopsis (Alonso-Blanco et al., 1999; Joosen et al., 2012), rice (Yoon et al., 2006; Shomura et al., 2008), soybean (Liu et al., 2007) and sunflower (Burke et al., 2002; Al-Chaarani et al., 2004). However, these studies provide little understanding of the developmental and molecular process regulating seed size variation. Tomato is one of the few species, where

comprehensive QTL mapping for seed weight has been conducted. Over the past 28 years, quantitative trait mapping studies, involving crosses between cultivated tomato and related wild species, have revealed many QTLs which account for most seed weight variation (Tanksley et al., 1982; Weller et al., 1988; Doganlar et al., 2000) including our current study (Khan et al., 2012), which, in addition to seed weight, also includes QTLs for seed dimensions (seed size, length and circularity).

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, 2006). The environment exerts great influence on seed size, with many factors that interact to affect the trait (Horii et al., 2006). Tomato seeds are composed of an embryo, an endosperm and the seed coat. Each of these three structures is genetically distinct. The embryo develops from the fertilized ovule and contains an equal representation of the maternal and paternal genomes, whereas the endosperm is usually formed by the fusion of two polar nuclei and one sperm nucleus and, therefore, contains two doses of the maternal parent's genes and one dose of the paternal parent's genes.

Vigour of seedlings immediately after germination is essential for good, sustainable and profitable crop production and seedling establishment is therefore considered the most critical stage of a crop. The effects of seed vigour on the emergence of seedlings and subsequent stand establishment are well documented (Roberts, 1972; Heydecker, 1977; TeKrony and Egli, 1991). Seedling vigour can potentially influence dry matter accumulation by the plant or plant community and thus immensely affect final yield of a crop. Poor seed vigour greatly influences both the number of seedlings that emerge, as well as the timing and uniformity of seedling emergence in all crops. This may have a major impact upon many aspects of crop production that determine cost effectiveness and the inputs required, and could also have direct influence on the marketing quality of a crop (Finch-Savage, 1995). Inadequate seedling growth will reduce total crop yield at harvest (Bleasdale, 1967) and no subsequent efforts or amount of inputs during later stages of crop development will compensate for this upshot. Abnormality at the time of seedling emergence can also affect the uniformity in plant size at harvest, which reduces the proportion of the crop in high-value size grades (Benjamin, 1990). In such a case the gross production may be high but the net profit of the crop can be greatly reduced due to low marketable yield. Seed vigour is therefore an important key factor which not only contributes directly to the economic success of commercial crops, but can also contribute in a number of indirect ways (Finch-Savage, 1995). For example, timing and uniformity of seedling emergence has an immediate impact upon the efficacy of herbicide applications, weeding strategies and other aspects of crop production that determine cost effectiveness. Poor seed quality also has a direct financial penalty for the production of transplants for vegetables and ornamentals in the glasshouse through wasted space, materials and reduced product quality resulting from non-uniformity.

Seed size is frequently measured as weight or volume, and, being an important component of seed quality, has a potential impact on seedling quality in many crop species (Wood et al., 1977; Rao, 1981; TeKrony and Egli, 1991). Generally, large seeds have better field performance than small seeds. Intra- and interspecific studies of offspring fitness in plant communities have demonstrated that plants producing large seeds often have higher tolerance to drought (Leishman and Westoby, 1994), herbivory (Bonfil, 1998), shading (Hewitt, 1998), and nutrient deficient soils (Jurado and Westoby, 1992). However, plants producing a large number of small seeds exhibit superior colonization abilities with the advantage of dispersal due to the abundance of seeds and higher likelihood to escape from predation (Coomes and Grubb, 2003; Gómez, 2004).

There is experimental evidence that larger seeds are better able to establish or survive as seedling in a variety of environments, including competition from established vegetation (Gross and Werner, 1982; Gross, 1984; Reader, 1993), competition from other seedlings (Black, 1958), drought (Wulff, 1986; Buckley, 1992), shading (Leishman and Westoby, 1994), mineral nutrient shortage (Lee and Fenner, 1989; Jurado and Westoby, 1992), and being covered by deeper or by little soil (Gulmon, 1992; Peterson and Facelli, 1992; Vázquez-Yanes and Orozco-Segovia, 1992). Although empirical evidence indicates that large seeds are beneficial only under some conditions, theoretical explanations for the maintenance of diversity of seed size have thus far focused exclusively on average performance, without considering habitat variation.

In cereal crops such as spring and winter wheat, (*Triticum aestivum* L.) seed size positively affected seedling establishment, shoot weight, forage production as well as grain yield under normal growing condition (Bockus and Shroyer, 1996). However, this effect becomes more pronounced under stress conditions (Mian and Nafziger, 1994). In soybean, individual seed weight and seedling growth rate were strongly correlated under high temperature stress (Dornbos Jr and Mullen, 1991) and the seedling from larger sized soybean varieties exhibited superior emergence, and vigourous seedling growth under both laboratory and field conditions (Burris et al., 1973). In addition to correlation between seed weight and seedling vigour traits, co-location of QTLs for these traits have been detected in several genetic studies for various species (Alonso-Blanco et al., 1999; Cui et al., 2002; Kehui et al., 2002; Groos et al., 2003; Burstin et al., 2007; Bettey et al., 2008; Finch-Savage et al., 2010), suggesting a common genetic mechanism underlying seed weight and seedling growth in different species.

The root system of a plant performs an essential role by 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 crucial role in determining whether a plant will survive in a particular environment (Malamy and Benfey, 1997). Heavy seed may have a better root architecture and seed size appears to have an essential role in an increased downward growth rate during its initial stage of

seedling growth (Jurado and Westoby, 1992). Dissecting the natural variation in seed vigour of *Brassica oleracea*, revealed a strong effect of seed vigour on the initial downward growth of seedlings and the co-locating QTLs for seed weight and rapid initial growth of the root have been fine mapped (Finch-Savage et al., 2010). In tomato, seed germination and early seedling growth are very sensitive stages to environmental stresses such as salinity, drought and extreme temperatures (Jones, 1986; Foolad et al., 2001). However, little is known about the role of tomato seed size in seedling vigour and establishment. No previous systematic genetic information is available about this aspect of seed quality. In the present study, as a result of extensive phenotyping of seed and seedling traits, seed reserves and metabolites, we have documented a strong genetic and physiological association among different seed dimensions and seedling vigour related traits. We show that seed dimensions in tomato such as size, weight and length have strong correlations with seedling traits and that there is co-location of QTLs for seed and seedling traits.

Seed quality and seed germination

In tomato, seed germination is the most sensitive stage of plant life that is greatly influenced by various environmental stresses including salt, temperature and water loss (Foolad et al., 1997; Foolad and Chen, 1999; Foolad et al., 2003; Foolad, 2007). These stresses may delay the onset, rate and uniformity of germination. Nevertheless, the impact of the environment depends to a large extent on the interaction between the genetic makeup of the plant and the environment and it is believed that the plant's response to environmental stresses is controlled by many genes (Foolad, 2007).

Completion of germination is defined as the protrusion of the radicle through the endosperm and seed coat (Bewley et al., 2012). During imbibition the embryonic axis elongates and breaks through the testa. Although seed size and/or weight is beneficial for seedling establishment and vigour related traits, there appears to be no consistent association between seed mass and seed germination performance (Fenner, 2006; Kazmi et al., 2012; Khan et al., 2012). Seed germination rather depends on the composition of seed reserves and the balance among different hormones and particularly abscisic acid (ABA)- and gibberellic acid (GA)-signalling that underpins germination potential, rather than one or the other alone (Penfield and King, 2009). Although recent studies on seed development have been invaluable in revealing aspects of the regulation of metabolism, investigation of the genetic basis of seed germination variability still remains 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 stress conditions. 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. Post-genomic technologies, such as transcriptomics, proteomics 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). Until now systematic studies to address the issue of seed quality in a multidisciplinary way have been lacking. The current study integrates different approaches to explore the underlying genetic and physiological causes regulating the complex traits relating to seed germination and seedling growth. This study is focusing on the systematic exploitation of the naturally occurring variation in tomato Recombinant Inbred Lines (RILs) obtained from a cross between *Solanum lycopersicum (*cv. Moneymaker) and *Solanum pimpinellifolium* (G1.1554) to provide new ways of dissecting the genetics of seed quality by combining the physiology, genetics and genomics to identify loci and genes that are responsible for seed quality traits.

Seed quality and seed reserves

Seed quality traits, such as seed germination and vigour, 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. The major storage compounds found in most mature seeds are proteins (mainly albumins, globulins, and prolamins), oil (often triacylglycerols) and carbohydrates (often starch) that are synthesized during the maturation phase of seed development (Baud et al., 2002; Bewley et al., 2012). The food reserves that seeds accumulate during the seed filling phase should provide sufficient nutrition and energy to the embryo during seed germination and early seedling growth. These reserves are of major importance as they support early seedling growth when degraded upon germination and, therefore, participate in crop establishment. The success of establishment and vigour of the young seedlings is determined by the quality of the seed and its interaction with the environment and the food reserves it contains are available to sustain the seedling until it becomes an independent, autotrophic organism, able to use light energy.

The duration of the seed filling phase and environmental conditions may potentially affect the amount and quality of reserve food stored. Thus, the seed filling phase indirectly plays a vital 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). These reserves may be stored in the different tissues of the seeds, depending on the species. For example, in dicots most of the reserves are located within the embryo tissues, including radicle, hypocotyl and, particularly, the cotyledons, whereas in monocots most of the storage reserves are accumulated in the endosperm (Bewley et al., 2012). Dicots such as legumes generally store higher levels of protein (21-40%) and oil as compared to starch. On the other hand most monocot seeds contain higher levels of

starch, located mainly in the endosperm and low levels of both protein and oil (Bewley et al., 2012). Tomato, being a dicot, contains high levels of protein (22-33%) as well as lipids (20-29%) and minor levels (0.5-2%) of starch (Schauer et al., 2005; Sheoran et al., 2005). Both the quality and quantity of the storage reserves is considerably influenced by the prevailing environmental conditions and the availability of carbon and nitrogen to the parent plant before and during their synthesis. Accumulation of starch and protein content in the seeds increases with the increase in concentration of nitrogen and carbon in the medium (Singletary and Below, 1989). In particular, the genotype and its interaction with the environment is an important attribute regulating the acquisition as well as composition of seed reserve food in a given genotype (Ries and Everson, 1973). Seed quality, among the other attributes, mainly depends on the amount and composition of protein, starch and oil, which are frequently defined as complex traits and are functionally dependent on the C-N balance, central metabolism and sink-source interaction during development on the mother plant (Wobus and Weber, 1999; Toubiana et al., 2012).

Despite the variety of seed storage products, the synthesis of all of these biopolymers utilizes sucrose, imported into the seeds from photosynthetic organs of the plant. Thus, it may be argued that the mechanism and regulation of carbon partitioning in seeds during development and maturation is integral to seed quality. The ultimate composition of the seed's food reserves depends on the relative sink strengths of the synthetic pathways of each individual reserve compound, as well as the sink strength of the diverse compartments where synthesis takes place, e.g. endosperm vs. embryo. Activities of key genes (and their products) of carbohydrate partitioning and conversion will be main determinants of the eventual composition of the food reserves. The sequences of most of these genes are known in Arabidopsis and may be used in gene expression studies during seed development, as well as in reverse genetics.

The relationship between seed performance and the amount of reserve food and its composition has so far received little attention in the seed literature (Castro et al., 2006). Seed vigour is a seed quality attribute, indicating the degree of stress tolerance of germination and seedling establishment. Seed storability and desiccation tolerance are acquired during the seed maturation phase, concomitantly with an increase in seed reserve and seed vigour. It is generally assumed that the increase in stress tolerance during seed development is a direct function of the accumulation of protecting proteins, including late embryogenesis abundant (LEA) proteins, peroxiredoxins and heat shock proteins (HSPs) (Delseny et al., 2001). Dormancy is a seed quality attribute that negatively affects both total germination and germination rate (Hilhorst and Koornneef, 2007). The acquisition of dormancy is commonly associated with the transient increase in content of ABA during seed development. In most of the species studied, ABA levels increase during the first half of seed development and decline during late maturation, concomitantly with the decline in seed water content. ABA has been detected in all seed and fruit tissues examined and has

been related to a number of developmental processes, including synthesis of storage proteins and late embryogenesis-abundant proteins, suppression of precocious germination, and induction of desiccation tolerance (Finkelstein et al., 2002; Koornneef et al., 2002; Hilhorst and Koornneef, 2007). Sensitivity to ABA plays an equally important role as ABA content in the induction of dormancy. The ABA-insensitive *abi1*, *abi2*, and *abi3* Arabidopsis mutants display variable reductions in seed dormancy (Koornneef et al., 1984). In addition, in the *abi3* mutant, also desiccation tolerance, degradation of chlorophyll and accumulation of storage compounds are abolished and *abi3* seeds display a poor longevity (Koornneef et al., 1984; Léon-Kloosterziel et al., 1996; Zeng et al., 2003). Thus, ABI3 (and other B3 type transcription factors, LEC2 and FUS3) are key elements in the regulation of seed development and maturation and, hence, may control such seed quality attributes as dormancy, vigour and storability. Furthermore, these transcription factors also regulate the seed storage protein genes *At2S3*, and *CRC* (cruciferin) which links storage protein accumulation to the acquisition of seed quality (Kroj et al., 2003). Thus, identification and functional classification of genes acting downstream in the ABA-signaling pathway(s) may yield valuable markers or modifiers of seed quality.

Seed reserve food, frequently represented by seed mass, potentially contributes to seedling vigour as it is generally assumed that larger seeds produce more vigorous seedlings (Poorter and Rose, 2005). Thus seed reserve food is considered to be an important attribute of the successful establishment and survival of seedlings. Seed size is often positively correlated with seed protein content and, in turn, seed protein content is frequently positively correlated with seedling vigour (Lowe and Ries, 1973; Ries and Everson, 1973; Evans and Bhatt, 1977; Saxena et al., 1987; Panthee et al., 2005). This suggests that large and heavy seeds will have a higher relative and total amount of protein and will produce more vigorous seedlings. In contrast, seed starch content is inconsistently correlated with seed or seedling mass. Most studies have revealed no or negative correlations, with the exception of a few in which grain starch content is positively correlated with grain weight and seedling biomass (Lai and McKersie, 1994; Cui et al., 2002; Sulpice et al., 2009; Ruffel et al., 2010). The genetic regulation of reserve food accumulation and seed and seedling biomass have been documented in several genetic studies and co-location of QTLs for seed reserves and seed and seedling traits have been identified (Cui et al., 2002; Groos et al., 2003; Burstin et al., 2007).

In tomato the endosperm serves as a source of food for the embryo during development and germination and the testa protects the embryo and endosperm in various environments. The genetic balance and interaction between the endosperm, embryo and maternal tissues is a basic requirement for normal seed development and remains one of the most complex and unresolved issue of seed development. Though embryo and endosperm are closely related seed components, yet they differentially correlated with seed weight and seedling vigour related traits in different crop species

(Zhang and Maun, 1993) and distinct accumulation of storage reserves has been documented in these two tissues of the seed (Singletary and Below, 1989; Lai and McKersie, 1994). Although numerous studies have shown the association between embryo and endosperm and their relation with seed and seedling quality phenotypes in food crops (López-Castañeda et al., 1996; Richards and Lukacs, 2002), little is known about the relationship between embryo and endosperm and their role in seed and seedling quality related traits in tomato. Therefore, the genetic dissection of seed processes regulating seed mass (reserve food) through molecular markers and QTL analysis, and their association with seed and seedling quality phenotypes, will contribute to unravelling the signalling pathways involved and will provide a means to predict and improve seed quality. Natural variation for seed reserve related traits existing in a RIL population is a valuable resource to unravel the complex genetic mechanisms involved in the acquisition of seed quality(Ligterink et al., 2012).

The Genetic Analysis of Natural Variation in Tomato

Intra-species genetic variation in morphology, physiology and environmental responses is universal. Natural variation provides the genetic material for natural selection and breeding ('artificial selection'). Genetic variation in nature often takes the form of a quantitative phenotypic range, with an approximately normal distribution, rather than of qualitative phenotypes that fall into discrete categories (Paran and Zamir, 2003). The classification of gene functions requires the phenotypic characterization of genetic variation. Currently, such functional characterization of genes is mainly based on analysis of laboratory-induced mutants that are selected in forward and reverse genetic studies. The naturally occurring genetic variation among different accessions is an alternative complementary source of genetic variation.

However, exploitation of the genetic variation among accessions has been limited because of its mostly quantitative nature, in contrast with the commonly studied mutants, which provide qualitative variation (Alonso-Blanco and Koornneef, 2000). Differences exist in the number of loci and the environmental effects influencing the variation under study, which determine the tools used for its analysis. Nevertheless, over the past decade the advent of efficient genetic methods to map quantitative trait loci (QTL) in combination with molecular marker technologies and specific statistical methods, which has established the map position and the effects of quantitative trait loci, allow this variation to be exploited up to the molecular level (Tanksley, 1993; Foolad and Chen, 1999; Alonso-Blanco and Koornneef, 2000; Mackay et al., 2009). There has been an increasing interest in exploring the natural variation among tomato accessions. Several studies have exploited natural variation to address 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). There are several ways to exploit natural variation, but central to the entire discipline of quantitative genetics is the concept of crosses among various accessions having distinct characters for the trait of interest (Alonso-Blanco and Koornneef, 2000). The resultant progenies derived therefrom segregate for a number of genetic traits and can be analyzed genetically for quantitative traits (Keurentjes et al., 2008). In this type of analysis the association of trait phenotypes with the genotype assayed by molecular markers is very effective for the analysis of QTLs, whereby the QTLs represent the genomic regions containing a locus or several closely linked loci, and their contribution to the total variance of the trait in that experiment. In plants the use of RIL mapping populations consisting of homozygous RILs is an important component of QTL analysis, and plays a key role in obtaining trait values from different replications and experiments performed under different environmental conditions. This kind of populations are obtained by single-seed descent from F_2 plants until F_6 or further generation(s) until the RILs become mostly homozygous. These populations are of great importance, as they are immortal and therefor a large 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. Sufficient natural variation and the complex nature of the traits of seed and seedling quality makes them suitable traits to decipher with a QTL approach.

Substantial natural variation for abiotic stress tolerance exists within cultivated tomato (*Solanum lycopersicum*), as well as in its related wild species such as *S. habrochaites, S. pimpinellifolium*, and *S. pennellii* (Wudiri and Henderson, 1985; Scott and Jones, 1986; Wolf et al., 1986). The wild type tomato germplasm is a rich source of desirable genetic variability as many wild species have been identified with high tolerance to both biotic and abiotic stresses (Rick, 1982). Among them *S. pimpinellifolium* offers several benefits for studying natural genetic variation and morphological characters. Phylogenetically, it is the most closely related wild species to *S. lycopersicum* and, hence, readily hybridized. Furthermore it is relatively well known genetically, amenable to experimental culturing, quickly growing, highly reproductive and relatively tolerant to biotic and abiotic stresses (Rick et al., 1977; Foolad et al., 2007). However, despite their close relationship, the two species have great natural variation in many morphologically and economically interesting traits, including fruit-, seed- and seedling quality related traits (Grandillo et al., 1999; Doganlar et al., 2000; Doganlar et al., 2002). In tomato, different QTLs for germination characteristics under stress (Foolad et al., 2003; Foolad et al., 2007; Kazmi et al., 2012) and for seed and seedling size (Doganlar et al., 2000; Khan et al., 2012) have been identified. In *Arabidopsis thaliana* different QTLs were found for dormancy

(Bentsink et al., 2010) and several germination characteristics (Clerkx et al., 2004; Galpaz and Reymond, 2010; Joosen et al., 2010; Joosen et al., 2012). In *Medicago truncatula* several QTLs were identified for germination at extreme temperatures (Dias et al., 2011) and germination and seedling growth under osmotic stress (Zeng et al., 2006; Vandecasteele et al., 2011; Vandecasteele et al., 2011; Vandecasteele et al., 2011). In rice, QTLs have been identified for seed storability (Zeng et al., 2006) and in lettuce QTLs have been detected for several germination characteristics, including thermoinhibition (Argyris et al., 2008). In spite of these and other studies on specific aspects of seed and seedling quality, a systematic study of the genetics of seed quality is still lacking. A more systematic approach studying genetic populations differing in seed and seedling quality phenotypes will provide valuable insight in the involvement of genes, and the processes they control, in the acquisition of seed quality. Until now, only a few QTL positions have been cloned and characterized in detail, but if genes or gene sets associated with seed quality parameters become available, they may be used as diagnostic tools to assess seed quality, in marker-assisted breeding, or in genetic modification to enhance seed quality.

Complex traits and generalized genetical genomics

Although phenotypic variation can be partly evaluated by examining one gene or mapping and characterizing loci that control a particular phenotype, this alone cannot fully explain the possible differences in the regulatory mechanisms of an organism due to the possible interaction among thousands of genes operating within most organisms (Phillips, 2008). Phenotypic traits are commonly known as complex traits, controlled by multiple genes, as well as environmental perturbations (Mackay, 2001; Phillips, 2008; Mackay et al., 2009). The phenotypic variation may occur due to variation at various molecular levels, such as variation in coding sequences; 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). For example, variation in coding sequences can alter protein function resulting in a changing metabolome in terms of chemical structure and function (Paran and Zamir, 2003). The recent shift towards integrating comprehensive functional genomics, and systems biology with high-resolution genetic mapping is now providing a more promising approach to address these issues more thoroughly than was possible in the past (Li et al., 2005; Phillips, 2008). This so called genetical genomics approach combines 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) However, one of the limitations of a standard genetical genomics approach is that it only takes effect of genetic perturbation for a single developmental stage or environment and usually does not take into account different environmental conditions. Since the complete understanding of most phenotypes requires studying them across different environments

and developmental stages, it is difficult to choose the most suitable developmental stage or environment. The current study seeks to resolve this issue by using a generalized-genetical-genomics (GGG) approach (Li et al., 2008) for tomato seed metabolomic analysis which takes into account both 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 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 unveils, for the first time, the plasticity of molecular networks in tomato for seed and seedling quality traits and forms a vital step toward understanding different influences of genetic to developmental and environmental responses of tomato seeds and seedling.

Transcriptomics and metabolomics for the dissection of complex traits

The rapid advances in 'omics' technologies have provided an opportunity to generate new datasets for crop species and have increased our understanding about multigenic traits, stress responses and defence mechanisms of higher plants (Langridge and Fleury, 2011). It is assumed that gene expression levels are affected by the functional polymorphisms that affect the trait of interest (Arbilly et al., 2006). Integration of genome and functional omics data with genetic and phenotypic information is leading to the identification of genes and pathways responsible for important agronomic phenotypes (Yuan et al., 2008). Transcriptomics, proteomics and, more recently, metabolomics are three of the most exciting new tools and techniques that are being used in all areas of biological research. When used in combination, they have the potential to comprehensively dissect a system at the transcriptional and translational level (Tan et al., 2009). Metabolomics is one of the most recent of these techniques to emerge and is concerned with the non-targeted profiling of all metabolites in a given biological system. In the genetical genomics strategy, the genetic mechanisms of segregation and recombination are used to reshuffle the genomes of two or more donor parents, to produce a population of segregating offspring (e.g. RILs, Introgression Lines (ILs) and Near Isogenic Lines (NILs)) with combinations of gene variants after which each individual of the population is used for genetic mapping and gene expression analysis (Brem et al., 2002). The expression level of each transcript in the segregating population is treated as a quantitative phenotype which is used to map loci affecting the gene expression levels, known as expression QTLs (eQTLs) (Jansen and Nap, 2001). Thus, the values of gene expression of all the individuals in a segregating population are used as a quantitative trait for QTL mapping. The effectiveness of the basic genetical genomics approach can be improved by carefully evaluating the possible experimental design (e.g. choosing the type of segregating population that is most suitable for

unravelling complex interactions) generating biological relevant models (such as those that take into account relevant biological or technical sources of variation) and the method of analysis (Jansen, 2003).

Genes are organized into regulatory circuits where the expression of one gene can influence the expression of another gene. Therefore, integrating observed expression profiles is not an easy task. A genetical genomics strategy is based on the idea that genes that function in the same pathway might have expression patterns that vary in the same way since they might be under the control of the same transcription factor. These genes are likely to map genetically to similar regions on the genome. This information helps in the construction of regulatory networks. Furthermore, eQTLs co-locating with the physical position of the gene on the genome (*cis*-acting genes) are considered good candidates for being the causal genes of functional quantitative trait loci (QTL) (Brem et al., 2002; Wayne and McIntyre, 2002). One of the first studies combining QTL analysis with gene expression profiling was carried out in yeast (Brem et al., 2002), shortly followed by maize (Schadt et al., 2003), eucalyptus (Kirst et al., 2004) and Arabidopsis (Keurentjes et al., 2007). 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).

Metabolomics is one of the more recent tools of crop analysis that are being applied for the sake of functional genomics. The ultimate goal of metabolomics is to be able to identify and measure a comprehensive profile of all, or at least as many as possible, different metabolites present in a biological sample (Verpoorte et al., 2008). Metabolites are quantitative in nature and a large and increasing body of literature has investigated the fact that metabolite abundance is generally regulated by multiple genes and metabolic QTLs (mQTL) (Kell et al., 2005; Lisec et al., 2007; Schauer et al., 2008; Toubiana et al., 2012). Metabolomics is often considered as a complementary technique to other functional genomics techniques (e.g. transcriptomics and proteomics). First, the metabolome more directly influences the phenotype than either transcripts or proteins do. Second, changes in the metabolome are often amplified relative to changes in the transcriptome or proteome (Sana et al., 2010). Experimental evidence based on investigation of the relationships of metabolites and developmental variations have established an integral link between plant central metabolism, growth and biomass accumulation (Keurentjes et al., 2007; Meyer et al., 2007). However, despite the strong association between metabolites and developmental traits, in several studies less than the expected association of metabolite QTL (mQTLs) with developmental traits has been reported. This lack of overlapping between known developmental and metabolic QTLs could be due to several reasons, including the size and structure of the mapping populations (Beavis, 1998; Clerkx et al., 2004; Rowe et al., 2008). In turn, this gives rise to the assumption that genetic regulation of plant metabolism is more complex than presumed, such that current studies resulted in

significantly higher number of phenotypic QTLs (phQTLs) as compared to metabolic (mQTLs).

Several successful studies have been conducted to date to identify novel genes based on QTL analysis (Kliebenstein et al., 2001; Kroymann et al., 2003; Werner et al., 2005; Zhang et al., 2006). In species such as Arabidopsis and tomato whose genomes are fully sequenced, identification of QTLs may provide a direct method for detection of functionally relevant variation in known genes with metabolic function and the identification of genes previously not assigned to metabolic functions, and may highlight the link between metabolism and growth/biomass accumulation. Such an example is a study in tomato, where the cause of a seed weight QTL has been associated with a gene encoding an ABC transporter gene by using genetic analysis (Orsi and Tanksley, 2009). Similarly Bentsink et al. (2010) have compared the dry seed transcriptomes of 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 good example of finding a causal gene by exploiting natural variation is the mapping of the Htg6.1 QTL in a lettuce RIL population for thermotolerance (Argyris et al., 2005; Argyris et al., 2008) which was further validated in NILs where it was subsequently confirmed to extend the range of germination under high temperature. Correlation analyses of shoot metabolites have revealed weak relationships between growth and the abundance of individual metabolites, but a close and highly significant link between biomass and a specific combination of metabolites has been shown (Meyer et al., 2007).

QTL confirmation and cloning

Detailed analysis of QTLs in segregating populations is limited by the resolution of QTL mapping which usually results in large chromosomal regions (Paterson et al., 1990). The capacity to map and manipulate genetic loci that condition the expression of a quantitative trait has blurred the distinction between the field of qualitative and quantitative genetics. Although considerable advancement has been made in fine mapping and cloning of genes underlying QTLs and reducing some of them to Quantitative Trait Nucleotides (QTNs), QTL mapping still remains a challenging task due to the large genetic intervals it produces, as well as QTLs of large effect which can be fragmented into several QTLs, explaining only a small proportion of the total variance. The dissection of quantitative traits using DNA markers has great potential both for improving the efficiency of plant breeding and for understanding and characterizing the physiological and biochemical processes associated with complex biological mechanisms (Dorweiler et al., 1993, Paterson et al., 1990). To obtain more precise map information, additional experiments are required. One approach to reduce the map position of a QTL is by analysing a series of near-isogenic lines (NILs) that differ for markers flanking the QTL of interest (Paterson et al., 1990; Kaeppler et al., 1993). With the help of this approach a small region of the genome that is

consistently associated with a quantitative trait and defines more precisely the map position for the QTL can be identified. Thus the NIL analysis allows identification of QTLs into smaller intervals as they differ in respect of overlapping regions of the genome indicated by QTL analysis (Tuinstra et al., 1997).

Although NILs are useful in the dissection of QTLs, this area of research has been limited by the cost, time, and effort required for developing the appropriate genetic materials (Tuinstra et al., 1997). Alternatively, NILs contrasting at the QTLs of interest can be developed by selection within heterogeneous inbred families (HIFs). HIFs are a set of lines derived from RILs that are genetically similar but have residual heterozygosity and still segregate for those loci that were heterozygous. A population of HIFs derived from different RILs can be screened through the use of molecular markers (Tuinstra et al., 1997). Thus the families that segregate for a specific region of the genome can be identified and a series of NILs that contrast for this specific region of the genome can be developed. This HIF approach is effective and less time consuming, as one does not need to develop the NILs first which is more time consuming and requires several generations of backcrossing and marker-assisted selection. Both NILs and HIFs can be used to confirm/validate the presence and effect of a QTL. An additional advantage of HIFs is their genomic composition which, although homozygous, is a mixture of the two distinct parental lines as compared to NILs which have a genetic background consisting of only one genotype (Loudet et al., 2005). The lines that reveal the predictable influence according to the QTL detection/validation should carry the gene that accounts to the effect of the QTL. Thus, a subset of RILs with residual heterozygosity can be used to developed HIFs families for further characterization and fine mapping of the QTLs of interest. This strategy can successfully be used for fine mapping in which lines with overlapping recombination events in the QTL region are phenotyped and the correlation between the phenotype and genotype thus narrow down the QTL interval.

Motivation of this Study

There is increasing interest in systematic characterization of the complex mechanisms regulating seed quality with respect to seed germination and early seedling growth. Most of these studies are based on QTL analysis and genetical genomics for searching regulatory genes which might govern complex networks and some of them have been successful in identifying causal genes controlling specific traits (Secko, 2005).

The main focus of genome research is on mapping and characterizing trait loci that control variation in various phenotypic characters that control growth, energy metabolism, development, reproduction and behaviour. These traits are generally known as complex traits, and are considered to have a multi-genic background governed by an unknown number of QTLs as well as many environmental perturbations (Andersson, 2001). Applying

the genetical genomics approach to embryo-derived tissue of germinating grains from the well-studied barley (*Hordeum vulgare*) *Steptoe* X *Morex (*St x Mx) segregating population, Kleinhofs and Han (2002) investigated the genetic control of gene expression. In the same population Potokina et al. (2008), identified 23,738 significant eQTLs affecting the expression of 12, 987 genes. They further observed that at least one eQTL hotspot was associated with at least one phenotypic phQTL for grain quality (such as grain protein content, alpha-amylase activity, diastatic power and malting quality) on different chromosomes. In a study using genetical genomics Kirst et al. (2004), assayed 2,608 genes in a backcross population of *E. grandis x E. globules* in Eucalyptus to reveal the genetic networks responsible for growth variation. They discovered two loci controlling lignin biosynthesis localized in the same genomic region as growth related QTLs. Therefore, it was suggested that the targeted regions regulate growth, lignin content and -composition. In Arabidopsis West et al. (2007) analyzed 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. It revealed 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). Some of such genetical genomics findings initially 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 and seedling quality phenotypes, a systematic way is needed to integrate and analyze the results generated by quantitative trait analyses, transcriptomic, metabolomic and seed reserves studies and molecular biological studies.

Thesis Objective

The objective of this thesis is to exploit the natural variation for seed and seedling quality related traits in tomato through molecular-genetic methods, tools and frameworks in order to obtain a better understanding of the mechanisms controlling these complex traits. The goal is to be able to characterize identified QTLs in the best possible way; (1) to explore which loci are likely to be responsible for a certain trait; (2) how these loci interact with each other; (3) what is the relationship between seed dimensions, seed reserve food, the level of seed metabolites and early seedling growth; (4) what is the proportion in which the environment (non-stress vs. stress) affects the phenotypic traits; (5) which loci have previously been reported in the same regions as the ones identified in the present study. This thesis makes efforts to get closer to the biological molecular-genetic interpretation of high-throughput data and the genetic characterization of QTLs by exploring and integrating

various sources of information, and ultimately target potential candidate genes that could be responsible for certain seed quality and seedling quality phenotypes.

Outline of the Thesis

This thesis consists of seven chapters including this general introduction (Chapter 1). Chapter 2 introduces the concepts of QTL mapping and looks at natural variation for seedling and root system architecture (RSA) traits and their link with seed dimensions present in a tomato RIL population. In addition to seed weight, one of the most significant aspects of this study is its emphasis on seed dimensions such as seed size. A strong relationship between different seed/seedling dimensions and RSA traits was established through phenotypic correlation and genetic co-location of QTLs, cementing the argument that larger seeds help in early growth and establishment of seedlings. Chapter 3 seeks for the genetic variation present in the RIL population that 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, environments and subsequent possibilities for improvement of tomato seed germination using selection. Chapter 4 explores the genetic variation present in the RIL population for two types of seed reserves, namely protein and starch content and their association with seed and seedling quality traits. A strong association between seed reserve and seed/seedling traits and RSA was found. Strong correlation of seed reserves and seed/seedling quality traits is supported by co-location of QTLs, supporting the concept that larger food reserves in large-sized seed helps in establishing more vigorous seedlings. Chapter 5 assesses the systems-genetics approach to find links between primary metabolites and seed quality and seedling 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. Chapter 6 demonstrates how the isolation of Heterogeneous Inbred Families (HIFs) helps with the confirmation/validation of QTLs. HIFs were constructed using the residual heterozygosity present in the F_8 lines of the *S. lycopersicum* x *S. pimpinellifolium* RIL population that allowed the unambiguous confirmation of QTLs for seed and seedling biomass on chromosomes 6 and 9 of the tomato genome. Chapter 7 discusses the main findings and overall contribution of the thesis and a final critical opinion about present and future research needed to follow up for a better understanding of complex seed and seedling quality traits.

References

Al-Chaarani GR, Gentzbittel L, Huang X, Sarrafi A (2004) Genotypic variation and identification of QTLs for agronomic traits, using AFLP and SSR markers in RILs of sunflower (*Helianthus annuus* L.). Theoretical and Applied Genetics **109:** 1353-1360

Alonso-Blanco C, Blankestijn-de Vries H, Hanhart C, Koornneef M (1999) Natural allelic variation at seed size loci in relation to other life history traits of *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences of the United States of America **96:** 4710

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

Andersson L (2001) Genetic dissection of phenotypic diversity in farm animals. Nature Reviews Genetics **2:** 130-138

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 Physiol **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. 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

Baud S, Boutin JP, Miquel M, Lepiniec L, Rochat C (2002) An integrated overview of seed development in *Arabidopsis thaliana* ecotype WS. Plant Physiology and Biochemistry **40:** 151-160

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

Benjamin L (1990) Variation in time of seedling emergence within populations: a feature that determines individual growth and development. Advances in Agronomy **44:** 1-25

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

Bettey M, Finch-Savage W, King G, Lynn J (2008) Quantitative genetic analysis of seed vigour and pre-emergence seedling growth traits in *Brassica oleracea*. New Phytologist **148:** 277-286

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

Bewley JD, Bradford KJ, Hilhorst HWM, Nonogaki H (2012) Seeds: Physiology of Development, Germination and Dormancy. Springer Verlag

Black C (1958) Soil-plant relationships. Soil Science **85:** 175

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

Bockus W, Shroyer J (1996) Effect of seed size on seedling vigor and forage production of winter wheat. Canadian Journal of Plant Science **76:** 101-105

Bonfil C (1998) The effects of seed size, cotyledon reserves, and herbivory on seedling survival and growth in *Quercus rugosa* and *Q. laurina* (Fagaceae). American Journal of Botany **85:** 79-79

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

Broich SL, Palmer RG (1980) A cluster analysis of wild and domesticated soybean phenotypes. Euphytica **29:** 23-32

Buckley J (1992) Universal fuzzy controllers. Automatica **28:** 1245-1248

Burke JM, Tang S, Knapp SJ, Rieseberg LH (2002) Genetic analysis of sunflower domestication. Genetics **161:** 1257-1267

Burris J, Edje O, Wahab A (1973) Effects of seed size on seedling performance in soybeans: II. Seedling growth and photosynthesis and field performance. Crop Science **13:** 207-210

Burstin J, Marget P, Huart M, Moessner A, Mangin B, Duchene C, Desprez B, Munier-Jolain N, Duc G (2007) Developmental genes have pleiotropic effects on plant morphology and source capacity, eventually impacting on seed protein content and productivity in pea. Plant Physiology **144:** 768-781

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

Clerkx E, Blankestijn-De Vries H, Ruys G, Groot S, Koornneef M (2004) Genetic differences in seed longevity of various Arabidopsis mutants. Physiologia Plantarum **121:** 448-461

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

Coomes DA, Grubb PJ (2003) Colonization, tolerance, competition and seed-size variation within functional groups. Trends in Ecology & Evolution **18:** 283-291

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

Cui K, Peng S, Xing Y, Xu C, Yu S, Zhang Q (2002) Molecular dissection of seedling-vigor and associated physiological traits in rice. Theoretical and Applied Genetics **105:** 745-753

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

Dias PM, Brunel-Muguet S, Durr C, Huguet T, Demilly D, Wagner MH, Teulat-Merah B (2011) QTL analysis of seed germination and pre-emergence growth at extreme temperatures in *Medicago truncatula*. Theoretical and applied genetics. Theoretische und angewandte Genetik **122:** 429-444

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

Dornbos Jr D, Mullen R (1991) Influence of stress during soybean seed fill on seed weight, germination, and seedling growth rate. Canadian Journal of Plant Science **71:** 373-383

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

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

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

Evans L, Bhatt G (1977) Iinfluence of seed size, protein content and cultivar on early seedling vigor in wheat Canadian Journal of Plant Science **57:** 929-935

Fenner M (2006) Relationships between seed weight, ash content and seedling growth in twenty-four species of Compositae. New Phytologist **95:** 697-706

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 WE, Clay HA, Lynn JR, Morris K (2010) Towards a genetic understanding of seed vigour in small-seeded crops using natural variation in *Brassica oleracea*. Plant Science **179:** 582-589

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

Finkelstein RR, Gampala SS, Rock CD (2002) Abscisic acid signaling in seeds and seedlings. Plant Cell **14:** S15-S45

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

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

Foolad M, Zhang L, Lin G (2001) Identification and validation of QTLs for salt tolerance during vegetative growth in tomato by selective genotyping. Genome **44:** 444-454

Foolad M, Zhang L, Subbiah P (2003) Genetics of drought tolerance during seed germination in tomato: inheritance and QTL mapping. Genome **46:** 536-545

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

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

Fuller DQ (2007) Contrasting patterns in crop domestication and domestication rates: recent archaeobotanical insights from the Old World. Annals of Botany **100:** 903-924

Galpaz N, Reymond M (2010) Natural variation in *Arabidopsis thaliana* revealed a genetic network controlling germination under salt stress. PLoS One **5:** e15198

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

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

Gómez JM (2004) Bigger is not always better: conflicting selective pressures on seed size in *Quercus ilex*. Evolution **58:** 71-80

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

Grandillo S, Ku H, Tanksley S (1999) Identifying the loci responsible for natural variation in fruit size and shape in tomato. Theoretical and Applied Genetics **99:** 978-987

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

Groos C, Robert N, Bervas E, Charmet G (2003) Genetic analysis of grain protein-content, grain yield and thousand-kernel weight in bread wheat. Theoretical and Applied Genetics **106:** 1032-1040

Gross K (1984) Effects of seed size and growth form on seedling establishment of six monocarpic perennial plants. The Journal of Ecology **72:** 369-387

Gross K, Werner P (1982) Colonizing abilities of'biennial'plant species in relation to ground cover: implications for their distributions in a successional sere. Ecology **63:** 921-931

Gulmon S (1992) Patterns of seed germination in Californian serpentine grassland species. Oecologia **89:** 27-31

Harlan JR, De Wet J, Price EG (1973) Comparative evolution of cereals. Evolution **27:** 311-325

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

Hewitt N (1998) Seed size and shade-tolerance: a comparative analysis of North American temperate trees. Oecologia **114:** 432-440

Heydecker W (1977) Stress and seed germination: an agronomic view. The physiology and biochemistry of seed dormancy and germination **237:** 282

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

Hilhorst HW, Finch-Savage WE, Buitink J, Bolingue W, Leubner-Metzger G (2010) Dormancy in plant seeds. *In* Dormancy and Resistance in Harsh Environments. Springer, pp 43-67

Horii H, Nemoto K, Miyamoto N, Harada J (2006) Quantitative trait loci for adventitious and lateral roots in rice. Plant Breeding **125:** 198-200

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

Huang X, Schmitt J, Dorn L, Griffith C, Effen S, Takao S, Koorneef M, Donohue K (2010) The earliest stages of adaptation in an experimental plant population: strong selection on QTLS for seed dormancy. Molecular Ecology **19:** 1335-1351

Isemura T, Kaga A, Konishi S, Ando T, Tomooka N, Han OK, Vaughan DA (2007) Genome dissection of traits related to domestication in Azuki bean (*Vigna angularis*) and comparison with other warm-season legumes. Annals of Botany **100:** 1053-1071

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

Janzen D (1969) Seed-eaters versus seed size, number, toxicity and dispersal. Evolution **23:** 1-27

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

Joosen RV, Arends D, Li Y, Willems LA, Keurentjes JJ, Ligterink W, Jansen RC, Hilhorst HW (2013) Identifying genotype-by-environment interactions in the metabolism of germinating Arabidopsis seeds using Generalized Genetical Genomics. Plant Physiology **162:** 553-566

Joosen RV, Kodde J, Willems LA, Ligterink W, van der Plas LH, Hilhorst HW (2010) Germinator: a software package for high-throughput scoring and curve fitting of Arabidopsis seed germination. 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

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

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

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

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

Kehui C, Shaobing P, Yongzhong X, Sibin Y, Caiguo X (2002) Molecular dissection of relationship between seedling characteristics and seed size in rice. Acta Botanica Sinica **44:** 702-707

Kell DB, Brown M, Davey HM, Dunn WB, Spasic I, Oliver SG (2005) Metabolic footprinting and systems biology: the medium is the message. Nature Reviews Microbiology **3:** 557-565

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 of the United States of America **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

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

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

Kleinhofs A, Han F (2002) Molecular mapping of the barley genome. Barley Science: Recent Advances from Molecular Biology to Agronomy of Yield and Quality**:** 31-63

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. Plant Cell **13:** 681-693

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

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

Koornneef M, Reuling G, Karssen C (1984) The isolation and characterization of abscisic acid-insensitive mutants of *Arabidopsis thaliana*. Physiologia Plantarum **61:** 377-383

Kroj T, Savino G, Valon C, Giraudat J, Parcy F (2003) Regulation of storage protein gene expression in Arabidopsis. Development **130:** 6065-6073

Kroymann J, Donnerhacke S, Schnabelrauch D, Mitchell-Olds T (2003) Evolutionary dynamics of an Arabidopsis insect resistance quantitative trait locus. Proceedings of the National Academy of Sciences of the United States of America **100:** 14587-14592

Lai FM, McKersie BD (1994) Regulation of starch and protein accumulation in alfalfa (*Medicago sativa* L.) somatic embryos. Plant Science **100:** 211-219

Langridge P, Fleury D (2011) Making the most of 'omics' for crop breeding. Trends in Biotechnology **29:** 33-40

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

Leishman M, Westoby M (1994) The role of large seed size in shaded conditions: experimental evidence. Functional Ecology **8:** 205-214

Léon-Kloosterziel KM, van de Bunt GA, Zeevaart JA, Koornneef M (1996) Arabidopsis mutants with a reduced seed dormancy. Plant Physiology **110:** 233-240

Li H, Lu L, Manly KF, Chesler EJ, Bao L, Wang J, Zhou M, Williams RW, Cui Y (2005) Inferring gene transcriptional modulatory relations: a genetical genomics approach. Human Molecular Genetics **14:** 1119-1125

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 RV, Hilhorst HW (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, Törjék O, Selbig J, Altmann T (2007) Identification of metabolic and biomass QTL in *Arabidopsis thaliana* in a parallel analysis of RIL and IL populations. Plant Journal **53:** 960-972

Liu B, Fujita T, Yan Z-H, Sakamoto S, Xu D, Abe J (2007) QTL mapping of domestication-related traits in soybean (*Glycine max*). Annals of Botany **100:** 1027-1038

López-Castañeda C, Richards R, Farquhar G, Williamson R (1996) Seed and seedling characteristics contributing to variation in early vigor among temperate cereals. Crop Science **36:** 1257-1266

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. Theoretical and Applied Genetics **110:** 742-753

Lowe L, Ries SK (1973) Endosperm protein of wheat seed as a determinant of seedling growth. Plant Physiology **51:** 57-60

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

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

Metz JA (1999) Evolutionary dynamics of seed size and seedling competitive ability. Theoretical Population Biology **55:** 324-343

Meyer RC, Steinfath M, Lisec J, Becher M, Witucka-Wall H, Törjék O, Fiehn O, Eckardt Ä, Willmitzer L, Selbig J (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

Mian M, Nafziger E (1994) Seed size and water potential effects on germination and seedling growth of winter wheat. Crop Science **34:** 169-171

Michaels HJ, Benner B, Hartgerink A, Lee T, Rice S, Willson MF, Bertin RI (1988) Seed size variation: magnitude, distribution, and ecological correlates. Evolutionary Ecology **2:** 157-166

Muller-Landau HC (2010) The tolerance–fecundity trade-off and the maintenance of diversity in seed size. Proceedings of the National Academy of Sciences of the United States of America **107:** 4242-4247

Nesi N, Delourme R, Brégeon M, Falentin C, Renard M (2008) Genetic and molecular approaches to improve nutritional value of *Brassica napus* L. seed. Comptes rendus biologies **331:** 763-771

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

Panthee D, Pantalone V, West D, Saxton A, Sams C (2005) Quantitative trait loci for seed protein and oil concentration, and seed size in soybean. Crop Science **45:** 2015-2022

Paran I, Zamir D (2003) Quantitative traits in plants: beyond the QTL. Trends in Genetics **19:** 303-306

Paterson A, DeVerna J, Lanini B, Tanksley S (1990) Fine mapping of quantitative trait loci using selected overlapping recombinant chromosomes, in an interspecies cross of tomato. Genetics **124:** 735

Penfield S, King J (2009) Towards a systems biology approach to understanding seed dormancy and germination. Proceedings of the Royal Society B: Biological Sciences **276:** 3561-3569

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

Peterson C, Facelli J (1992) Contrasting germination and seedling growth of *Betula alleghaniensis* and *Rhus typhina* subjected to various amounts and types of plant litter. American Journal of Botany **79:** 1209-1216

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

Poorter L, Rose SA (2005) Light-dependent changes in the relationship between seed mass and seedling traits: a meta-analysis for rain forest tree species. Oecologia **142:** 378-387

Potokina E, Druka A, Luo Z, Wise R, Waugh R, Kearsey M (2008) Gene expression quantitative trait locus analysis of 16.000 barley genes reveals a complex pattern of genome-wide transcriptional regulation. Plant Journal **53:** 90-101

Powell A, Basra A (2006) Seed vigor and its assessment. Handbook of seed science and technology**:** 603-648

Rao S (1981) Influence of seed size on field germination, seedling vigour yield and quality in self pollinated crops-a review. Agricultural Reviews **2**

Reader R (1993) Control of seedling emergence by ground cover and seed predation in relation to seed size for some old-field species. Journal of Ecology **81:** 169-175

Richards R, Lukacs Z (2002) Seedling vigour in wheat-sources of variation for genetic and agronomic improvement. Crop and Pasture Science **53:** 41-50

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 in *Lycopersicon pimpinellifolium*: Evidence of evolutionary change in mating systems. Plant Systematics and Evolution **127:** 139-170

Ries S, Everson E (1973) Protein content and seed size relationships with seedling vigor of wheat cultivars. Agronomy Journal **65:** 884-886

Roberts E (1972) Loss of viability and crop yields. *In* Viability of seeds. Springer, pp 307-320

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

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

Sana TR, Fischer S, Wohlgemuth G, Katrekar A, Jung K-h, Ronald PC, Fiehn O (2010) Metabolomic and transcriptomic analysis of the rice response to the bacterial blight pathogen *Xanthomonas oryzae* pv. oryzae. Metabolomics **6:** 451-465

Saxena K, Faris D, Singh U, Kumar R (1987) Relationship between seed size and protein content in newly developed high protein lines of pigeonpea. Plant Foods for Human Nutrition (Formerly Qualitas Plantarum) **36:** 335-340

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

Schauer N, Zamir D, Fernie AR (2005) Metabolic profiling of leaves and fruit of wild species tomato: a survey of the Solanum lycopersicum complex. Journal of Experimental Botany **56:** 297-307

Scott S, Jones R (1986) Cold tolerance in tomato. II. Early seedling growth of Lycopersicon spp. Physiologia Plantarum **66:** 659-663

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

Sheoran IS, Olson DJH, Ross ARS, Sawhney VK (2005) Proteome analysis of embryo and endosperm from germinating tomato seeds. Proteomics **5:** 3752-3764

Shomura A, Izawa T, Ebana K, Ebitani T, Kanegae H, Konishi S, Yano M (2008) Deletion in a gene associated with grain size increased yields during rice domestication. Nature genetics **40:** 1023-1028

Singletary GW, Below FE (1989) Growth and composition of maize kernels cultured in vitro with varying supplies of carbon and nitrogen. Plant Physiology **89:** 341-346

Smith CC, Fretwell SD (1974) The optimal balance between size and number of offspring. American Naturalist **108:** 499-506

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

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

Sulpice R, Pyl ET, Ishihara H, Trenkamp S, Steinfath M, Witucka-Wall H, Gibon Y, Usadel B, Poree F, Piques MC (2009) Starch as a major integrator in the regulation of plant growth. Proceedings of the National Academy of Sciences of the United States of America **106:** 10348-10353

Tan K, Ipcho SV, Trengove RD, Oliver R, P, Solomon P, S (2009) Assessing the impact of transcriptomics, proteomics and metabolomics on fungal phytopathology. Molecular plant pathology **10:** 703-715

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

Tanksley SD, Medina-Filho H, Rick CM (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

TeKrony DM, Egli DB (1991) Relationship of seed vigor to crop yield: a review. Crop Science **31:** 816-822

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. Theoretical and Applied Genetics **95:** 1005-1011

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

Vandecasteele C, Teulat-Merah B, Morere-Le Paven MC, Leprince O, Ly Vu B, Viau L, Ledroit L, Pelletier S, Payet N, Satour P, Lebras C, Gallardo K, Huguet T, Limami AM, Prosperi JM, Buitink J (2011) Quantitative trait loci analysis reveals a correlation between the ratio of sucrose/raffinose family oligosaccharides and seed vigour in *Medicago truncatula*. Plant Cell Environment **34:** 1473-1487

Vázquez-Yanes C, Orozco-Segovia A (1992) Effects of litter from a tropical rainforest on tree seed germination and establishment under controlled conditions. Tree Physiology **11:** 391

Verpoorte R, Choi Y, Mustafa N, Kim H (2008) Metabolomics: back to basics. Phytochemistry Reviews **7:** 525-537

Wayne ML, McIntyre LM (2002) Combining mapping and arraying: an approach to candidate gene identification. Proceedings of the National Academy of Sciences of the United States of America **99:** 14903-14906

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

Werner JD, Borevitz JO, Warthmann N, Trainer GT, Ecker JR, Chory J, Weigel D (2005) Quantitative trait locus mapping and DNA array hybridization identify an FLM deletion as a cause for natural flowering-time variation. Proceedings of the National Academy of Sciences of the United States of America **102:** 2460-2465

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

Westoby M, Jurado E, Leishman M (1992) Comparative evolutionary ecology of seed size. Trends in Ecology & Evolution **7:** 368-372

Westoby M, Leishman M, Lord J, Poorter H, Schoen DJ (1996) Comparative ecology of seed size and dispersal [and discussion]. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences **351:** 1309-1318

Willson M (1983) Plant reproductive ecology. New York

Wobus U, Weber H (1999) Seed maturation: genetic programmes and control signals. Current Opinion in Plant Biology **2:** 33-38

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, Longden P, Scott R (1977) Seed size variation; its extent, source and significance in field crops. Seed Science Technology **5:** 337-352

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

Yoon D-B, Kang K-H, Kim H-J, Ju H-G, Kwon S-J, Suh J-P, Jeong O-Y, Ahn S-N (2006) Mapping quantitative trait loci for yield components and morphological traits in an advanced backcross population between Oryza grandiglumis and the *O. sativa* japonica cultivar Hwaseongbyeo. Theoretical and Applied Genetics **112:** 1052-1062

Yuan JS, Galbraith DW, Dai SY, Griffin P, Stewart Jr CN (2008) Plant systems biology comes of age. Trends in Plant Science **13:** 165-171

Zeng D, Guo L, Xu Y, Yasukumi K, Zhu L, Qian Q (2006) QTL analysis of seed storability in rice. Plant Breeding **125:** 57-60

Zeng Y, Raimondi N, Kermode AR (2003) Role of an ABI3 homologue in dormancy maintenance of yellow-cedar seeds and in the activation of storage protein and Em gene promoters. Plant Molecular Biology **51:** 39-49

Zhang J, Maun M (1993) Components of seed mass and their relationships to seedling size in *Calamovilfa longifolia*. Canadian Journal of Botany **71:** 551-557

Zhang Z, Ober JA, Kliebenstein DJ (2006) The gene controlling the quantitative trait locus EPITHIOSPECIFIER MODIFIER1 alters glucosinolate hydrolysis and insect resistance in Arabidopsis. Plant Cell **18:** 1524-1536

Chapter 2

Natural Variation for Seedling Traits and their Link with Seed Dimensions in Tomato

Khan N*, Kazmi RH*, Willems LAJ, van Heusden AW, Ligterink W, Hilhorst HWM

Published in PLoS One 7: e43991 (2012)

*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 4π*(area/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 2.1). In addition, there were statistically significant differences for these traits among the different lines of the RIL population (Table 2.1).

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 2.1). 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 days after germination under the normal nutrient conditions (Table 2.2). All measured traits showed a normal distribution over the RIL population (Figure 2.1). Figure 2.1 also shows that transgression was present for most traits.

Correlation between traits

Statistically significant correlations were observed between seed weight and seedling traits such as fresh and dry root and shoot weight (Figure 2.2). 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 (Figure 2.2). 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

Table 2.1. 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 nutrientless 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.

Table 2.2. 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 nutrient conditions; [2]Wn= Seedling grown on Copenhagen table without nutrition; [3]Decr = Percentage decrease in biomass of seedling grown on Copenhagen table without nutrition compared to normal nutrient conditions; [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.

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 2.2). 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 2.2).

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 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 2.3). 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 (Figure 2.3). Co-localization of QTLs was found for different seed and seedling traits on the bottom of chromosomes 1, 4, 6, 9 and 11 (Table 2.3, Figure 2.3). Out of the 62 detected QTLs, 25 were related to seed traits, such as seed weight, seed size, seed length and seed circularity.

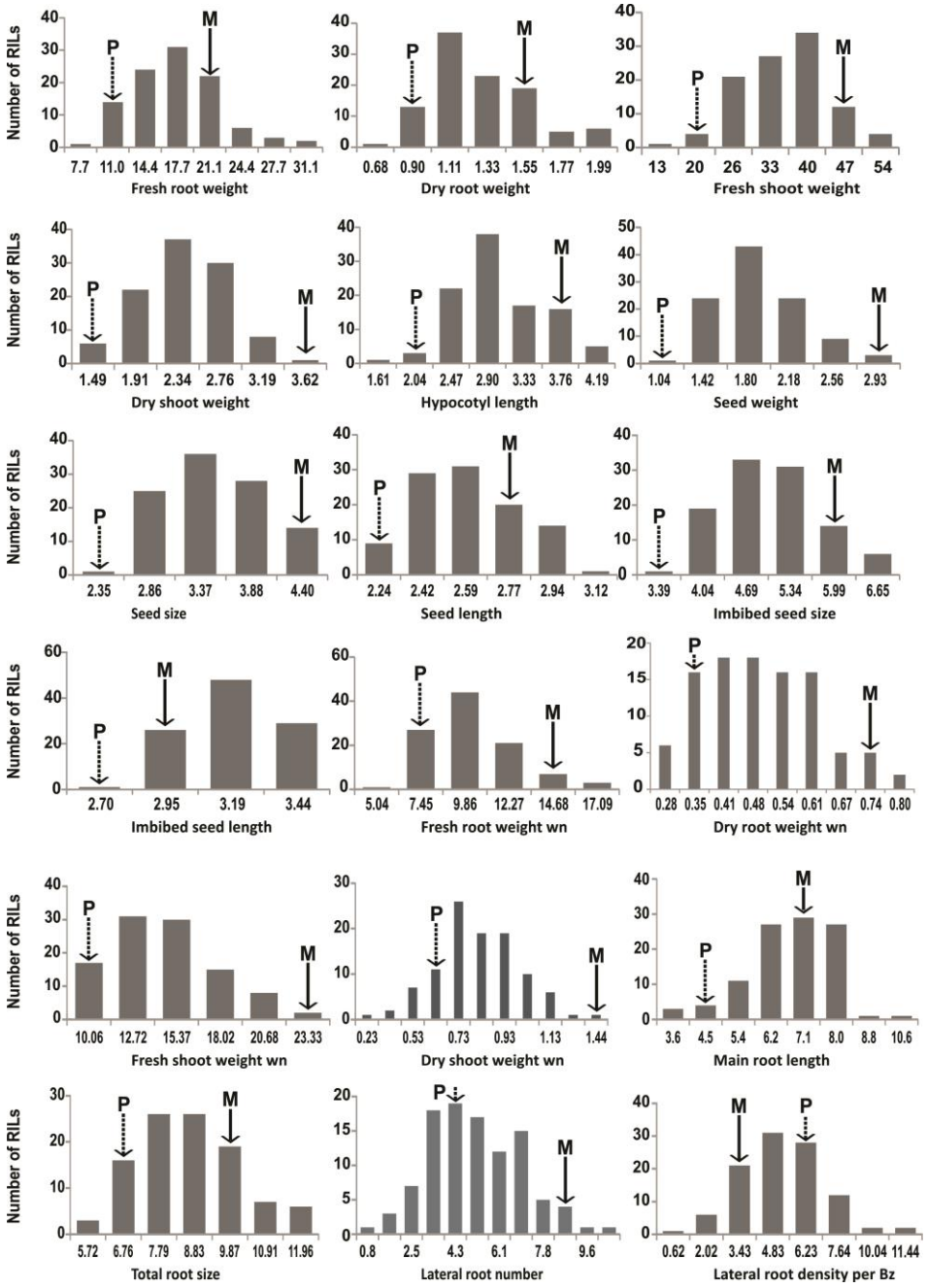


Figure 2.1. Frequency distributions of non-normalized data of all measured seed and seedling phenotypes in the *Solanum lycopersicum* x *Solanum pimpinellifolium* RIL population. wn: without nutrition. The parental values are indicated with a solid arrow. P = *S. pimpinellifolium* parent and M= *S. lycopersicum* parent.

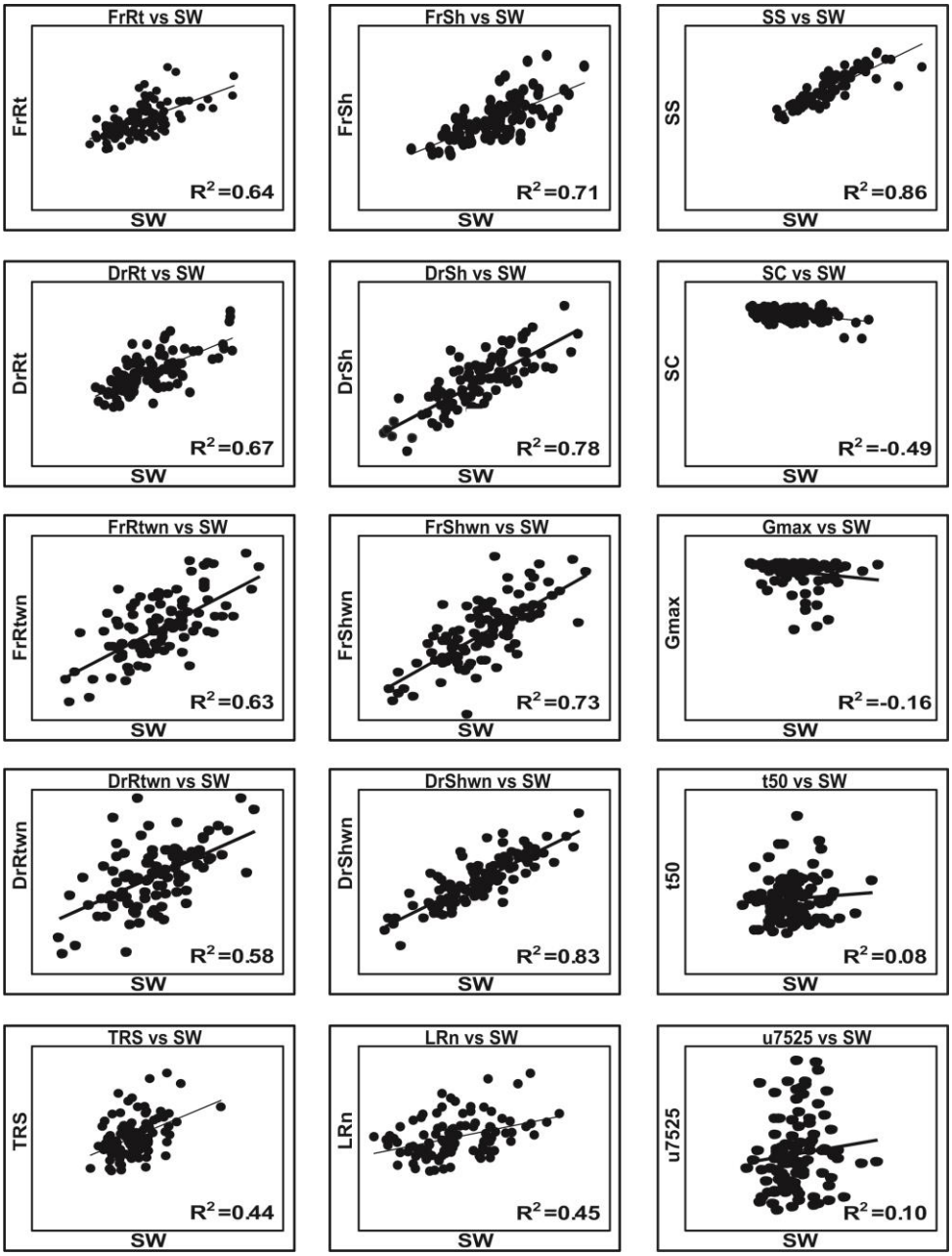


Figure 2.2. Correlation among seed and seedling traits. SW = Seed weight, SS = Seed size, SL = Seed length, FrRt = Fresh root weight, DrRt = Dry root weight, FrSh = Fresh shoot weight, DrSh = Dry shoot weight, FrShwn = Fresh shoot weight in nutrientless conditions, DrShwn = Dry shoot weight in nutrientless 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 2.3. Overview of significant QTLs associated with seed and seedling traits of *S. lycopersicum* and *S. pimpinellifolium* tomato RIL population.

Trait	Chr[2]	Confidence Interval (cM)	Nearest Marker[3]	LOD score	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		
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		
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		

TRS								
	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.

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 from

11.9 % 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 2.3).

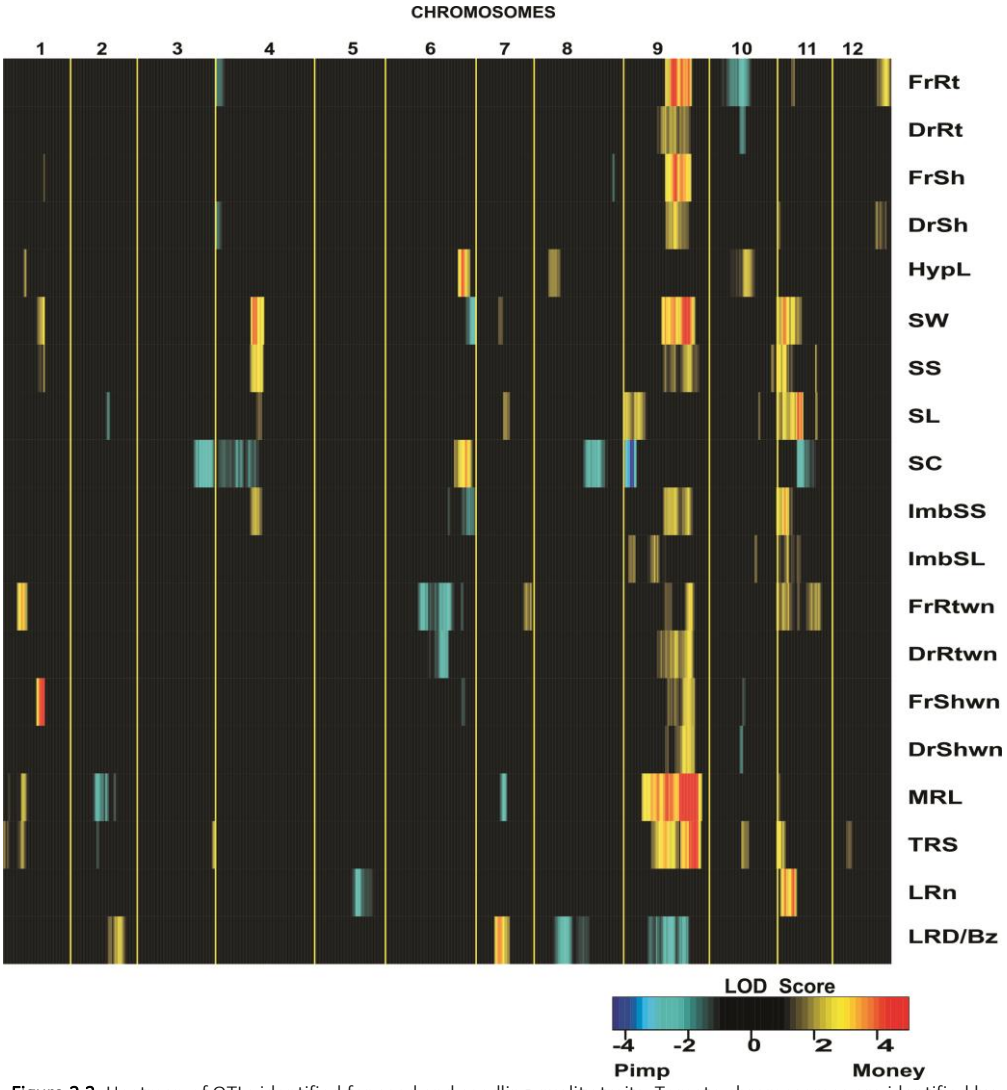


Figure 2.3. Heatmap of QTLs identified for seed and seedling quality traits. Tomato chromosomes are identified by numbers (1-12), with centimorgans ascending from the left to right; chromosomes are separated by yellow lines. SW=Seed Weight, SS=Seed Size SL=Seed Length. FrRt= Fresh Root weight, DrRt= 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 2.3, Figure 2.3). 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 2.4, Figure 2.4). 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 2.4, Figure 2.4). 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 2.4, Figure 2.4), which had the highest level of statistical significance obtained in our epistasis screen.

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.

Table 2.4. 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.

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 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.

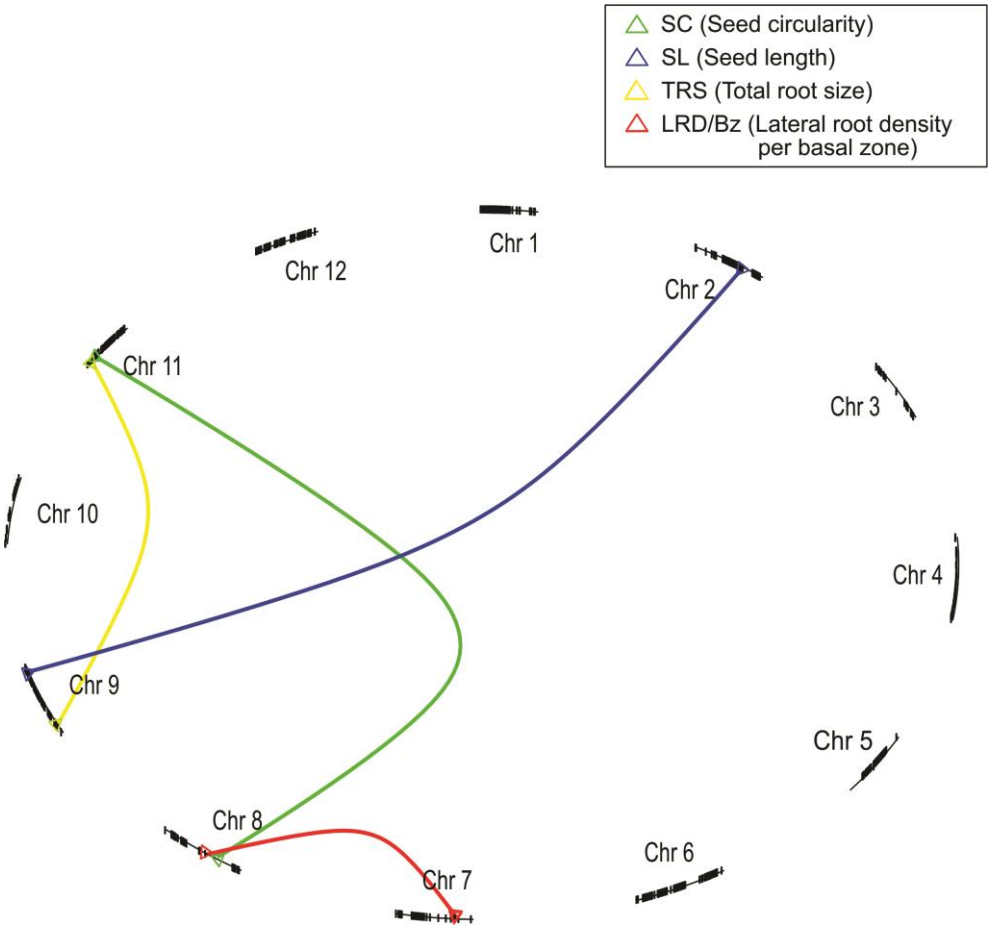


Figure 2.4. Epistatic interaction network of QTLs identified for seed and seedling quality traits. Graphical visualization of the epistatic interactions observed among different loci controlling seed and seedling quality phenotypes. The 12 chromosomes are represented as different circle segments, and 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).

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 2.3), 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 2.3).

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 (Figure 2.2). 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 2.3, Figure 2.3). 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 2.3 and Figure 2.3). 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. 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 **71:** 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 **179**: 582-589

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 **23:** 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. 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 in *Lycopersicon 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 Wild tomate *Lycopersicon pimpinellifolium* (Jusl.) Mill. I. Genetic Resources and Crop Evolution **8:** 110-137

Stubbe H (1965) Mutanten der Wild tomate *Lycopersicon 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 **57:** 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

Wilson M (1983) Plant reproductive ecology. Wiley, 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 3

Seed Quality Phenotypes in a Recombinant Inbred Population of an Interspecific Cross between *Solanum lycopersicum* x *Solanum pimpinellifolium*

Kazmi RH*, Khan N*, Willems LAJ, van Heusden AW, Ligterink W, Hilhorst HWM

Published in Plant Cell & Environment (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 3.1). 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.

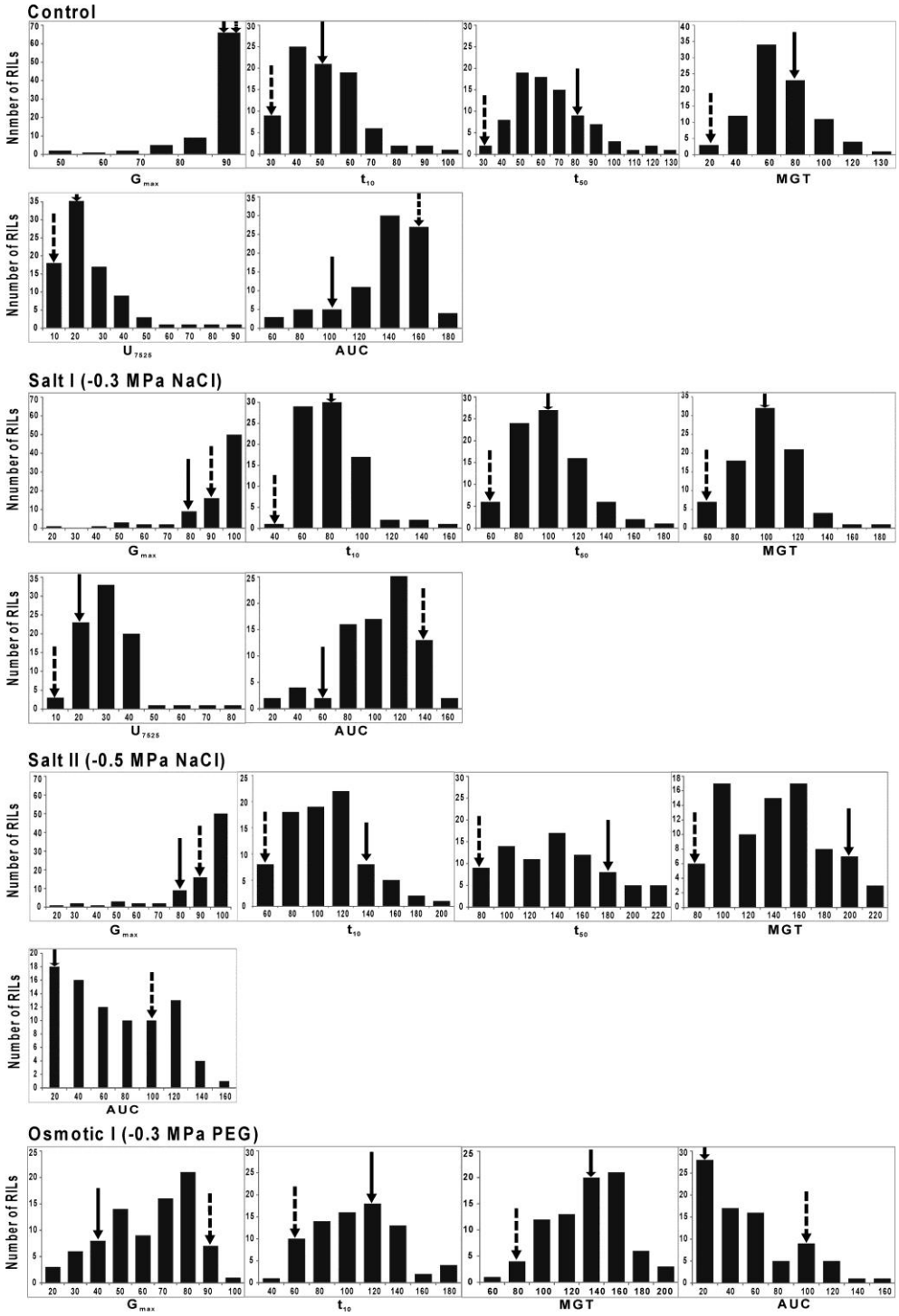
Table 3.1. Means of germination traits (±SD) for the parental genotypes and the F_8population of cross between *S. lycopersicum* (Money) and *S. pimpinellifolium* (Pimp) in the control- (non-stress), salt-, osmotic-, cold-, temperature- and oxidative-stress treatments.

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.

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 3.1, Figure 3.1). However, we also observed transgressive segregation for the seed quality traits (Table 3.1, Figure 3.1). 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 3.2). 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 3.1). 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 3.1).


Control
Nnmber of RILs
G_{max}
t_{10}
t_{50}
MGT
Nnumber of RILs
U_{7525}
AUC
Salt I (-0.3 MPa NaCl)
Nnumber of RILs
G_{max}
t_{10}
t_{50}
MGT
Number of RILs
U_{7525}
AUC
Salt II (-0.5 MPa NaCl)
Number of RILs
G_{max}
t_{10}
t_{50}
MGT
Number of RILs
AUC
Osmotic I (-0.3 MPa PEG)
Number of RILs
G_{max}
t_{10}
MGT
AUC

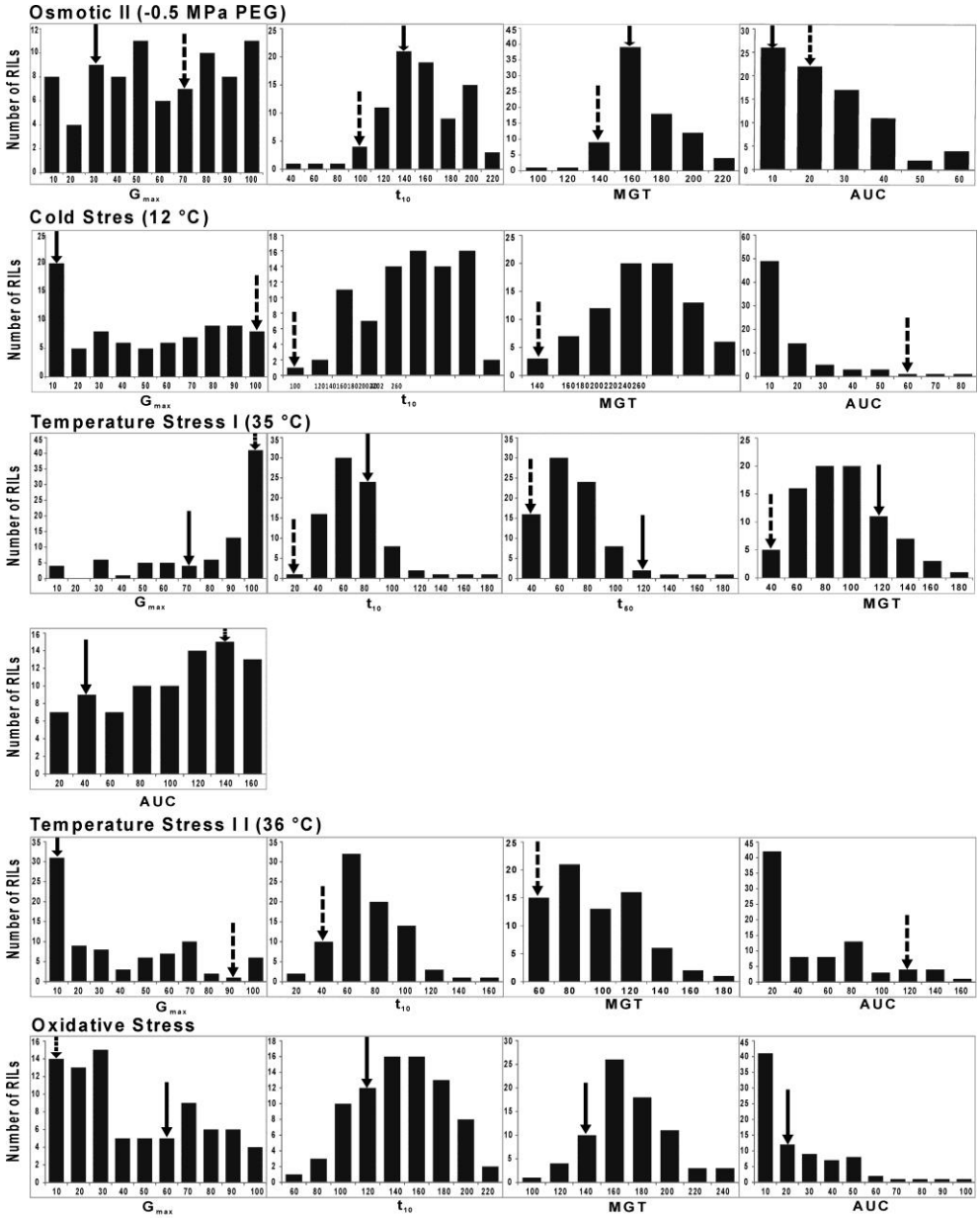


Figure 3.1. Frequency distributions of non-normalized data of all traits in the *Solanum lycopersicum* and *Solanum pimpinellifolium* recombinant inbred line (RIL) population. Seed quality traits determined under control conditions, salt stress I (-0.3 MPa NaCl), salt stress II (-0.5 MPa NaCl), osmotic 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 3.2 and 3.3. 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 3.2 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 3.2).

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 3.1, 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 3.2 and 3.3, Figure 3.2). 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 3.2 and 3.3, Figure 3.2). 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 3.2 and 3.3, Figure 3.2).

QTL for germination under salt stress conditions

Several QTLs were found to be associated with the tested germination traits (Tables 3.2 and 3.3) 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 3.2 and 3.3, Figure 3.2). 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 3.2 and 3.3, Figure 3.2). Furthermore, for U_{7525}^{-1} under low salt stress two QTLs were identified at chromosomes 4 and 7.

Table 3.2. 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 Interval (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	
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	
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
	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}								0.80
	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	
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_2O_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	
MGR								
	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								0.90
	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 $h^2_b = (\sigma^2 g / \sigma^2 g + \sigma^2 e)$.

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 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 4, 6 and 9 which explained 33.9% of variance (Table 3.2 and 3.3, Figure 3.2). 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 3.2).

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 3.2 and 3.3). 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 3.2 and 3.3, Figure 3.2). 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 3.2 and 3.3, Figure 3.2). 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 3.2 and 3.3, Figure 3.2), 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 the osmotic stress levels (Figure 3.2).

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 3.2 and 3.3, Figure 3.2). For t_{10}^{-1}, four QTL were found on chromosome 4, 5, 6 and 7 with 42.1% of total explained variance (Table 3.2 and 3.3, Figure 3.2), 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 3.2 and 3.3, Figure 3.2). 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 3.2 and 3.3, Figure 3.2). 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 total variance, respectively.

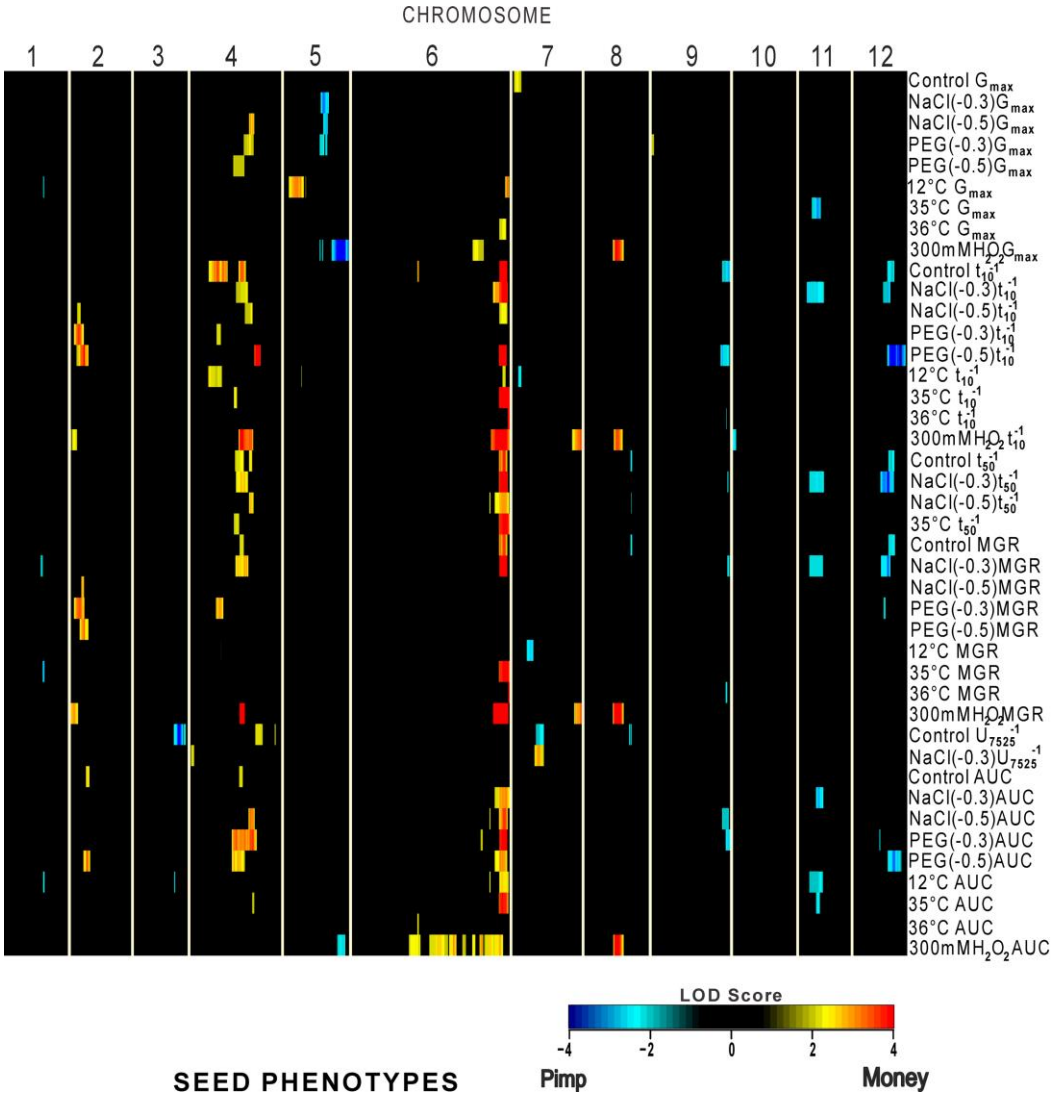


Figure 3.2. Genomic locations of quantitative trait locus (QTL) identified for seed quality traits. Tomato chromosomes are identified by numbers (1–12), with centimorgans ascending from the left to right; chromosomes are separated by white lines. Control indicates germination phenotypes under optimal 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 3.2. AUC, area under the germination curve; MGR, mean germination rate.

Table 3.3. 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
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 3.2 and 3.3, Figure 3.2).

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 3.2 and 3.3, Figure 3.2). 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 3.2 and 3.3, Figure 3.2). 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 3.2 and 3.3, Figure 3.2).

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 3.2). 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 3.2). The overlapping range between QTLs affecting simultaneously t_{10}^{-1}, t_{50}^{-1} and MGR varied from 90.0 to 100% (Figure 3.2). 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 3.2).

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 Supplemental Table S3.1. 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 3.3 and Supplemental Figure S3.1, Supplemental Table S3.1).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 3.3 and Supplemental Figure S3.1, Supplemental Table S3.1).

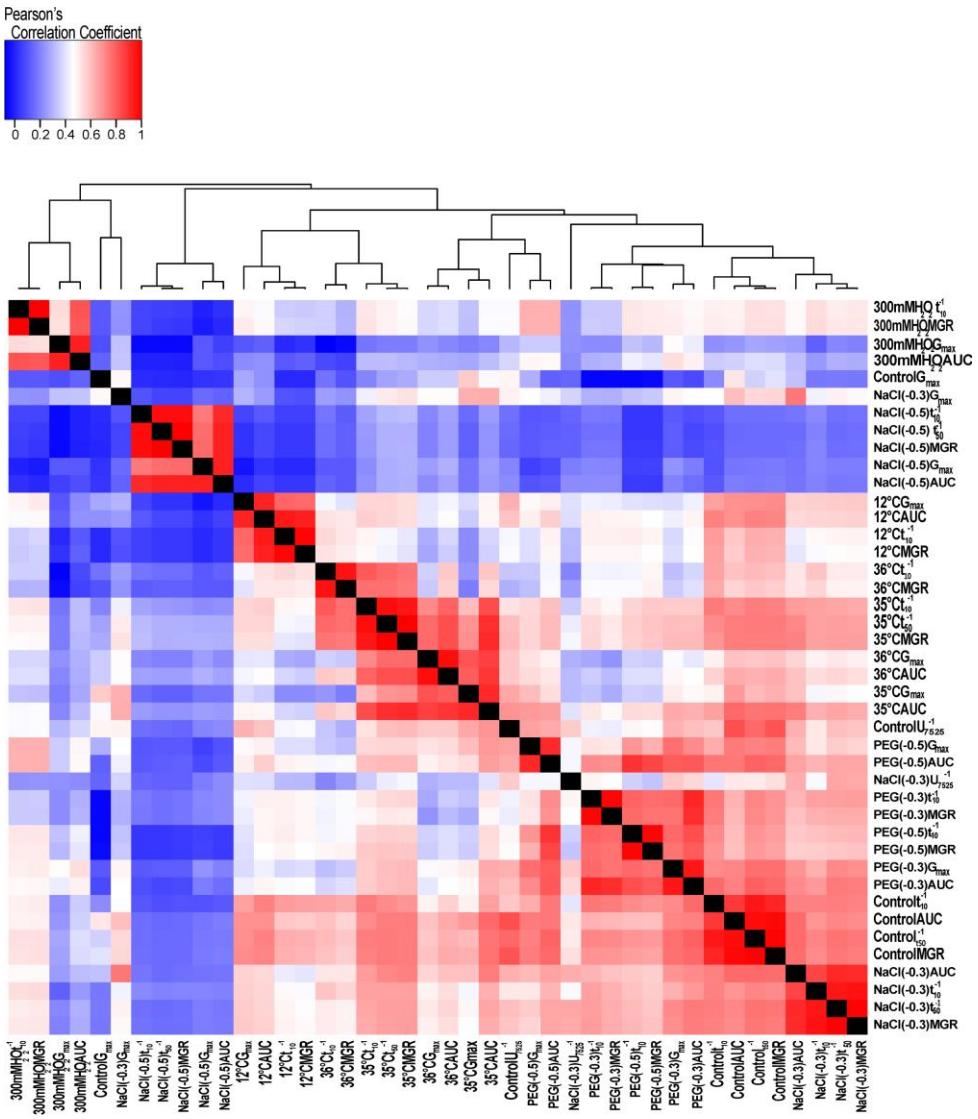


Figure 3.3. Heatmap of correlations between seed quality phenotypes. Each square represents the Pearson correlation coefficient between the seed phenotypes of the column with that of the row. Seed phenotype order is determined as in hierarchical clustering using the distance function 1-correlation. The 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 Supplemental Table S3.1. Supplemental Figure S3.1 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.

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 3.3 and Supplemental Figure S3.1, Supplemental Table S3.1).

Epistasis

The results of genome-wide epistasis analysis for each of the seed quality phenotypes are presented in Table 3.4. 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.

Table 3.4. 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.

The analysis revealed a locus on chromosomes 4 and 5 interacting to influence U_{7525}^{-1} under control conditions (Table 3.4, Figure 3.4). 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 3.4, Figure 3.4). An epistatic interaction was also observed for AUC under cold stress (12 °C) between QTL on chromosome 3 and 11 (Table 3.4, Figure 3.4).

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 3.2).

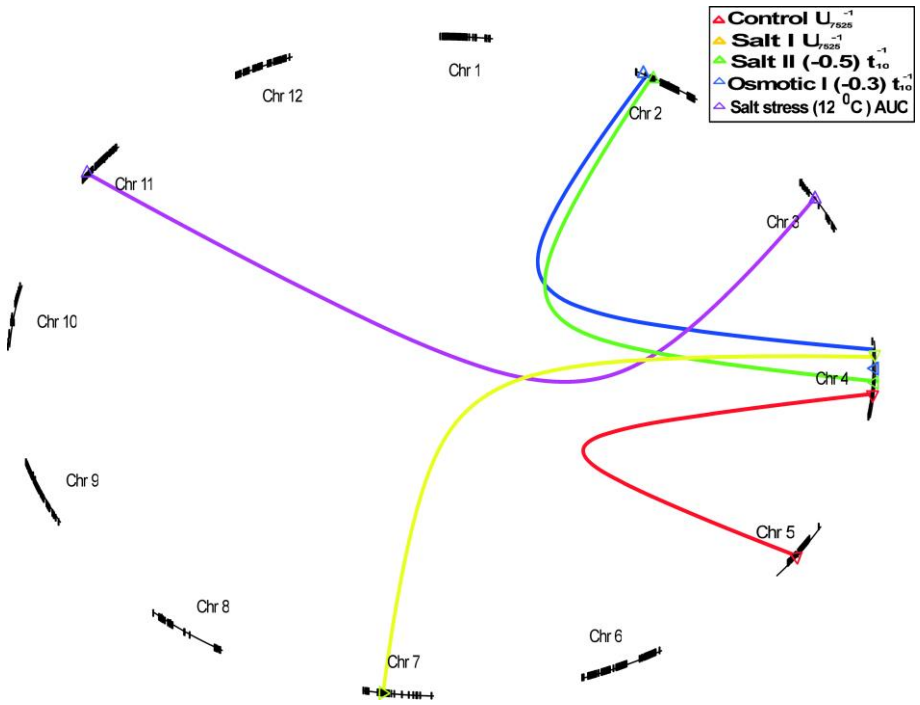


Figure 3.4. Epistatic interaction network. Graphical visualization of the epistatic interactions found between different loci controlling seed quality phenotypes in *Solanum lycopersicum* and *Solanum pimpinellifolium* recombinant inbred line population. The 12 chromosomes are represented as different 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.

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 3.2). Another instance of significant co-locations of QTLs was identified for these seed quality traits on linkage groups 1, 2, 9 and 12 (Figure 3.2). 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 3.2). 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 3.2). 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 3.2). 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 3.3). 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 3.2). 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.).

Supplemental Files

Supplemental can be downloaded from either the online version of this article (Kazmi et al. 2012) or from http://www.wageningenseedlab.nl/thesis/nkhan/SI/chapter3

Supplemental Table S3.1. Self-self correlations are identified in black. Individual correlation coefficients can be found in Table S3.1.

Supplemental Figure S3.1. 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 **29:** 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 4

Genetic Analysis of Whole Seed and Tissue-Specific Food Reserves Reveals a Close Link between the Abundance of Seed Reserves and Seed and Seedling Biomass.

Khan N, Willems LAJ, Ligterink W, Hilhorst HWM

Abstract

It is generally believed that large seeds establish vigorous seedlings and that seed and seedling vigour is due to a higher amount of reserve food in the larger seeds. We explored possible mechanisms involved in the accumulation of higher reserve food contents in the seed and its link with seed quality. To this end we have explored the genetic variation found in a RIL population of *Solanum lycopersicum* (cv. Moneymaker) x *Solanum pimpinellifolium* through extensive phenotyping of seed protein and starch content. We have identified 24 major QTLs on 16 different loci for both relative and total amount of seed protein and starch in this RIL population. Most of the QTLs identified for protein content co-located with seed and seedling QTLs that we had identified in a previous study and the positive alleles for all the protein QTLs were contributed by the *S. lycopersicum* parent, whereas only 4 out of the 17 QTLs identified for starch co-located with seed weight QTLs. Co-location of QTLs for different traits might indicate that the same locus has pleiotropic effects on multiple traits due to a common mechanistic basis. In an additional experiment we dissected seeds of the 20 most extreme lines of the RIL population, including the two parents and determined embryo and endosperm weight, protein and starch contents. We found highly significant correlations between, embryo-, endosperm- and seed weight. Further, we identified significantly higher levels of relative amounts of protein in the embryo than in the endosperm. We demonstrate here that the reserve protein of tomato seed makes an important contribution to seed weight compared to seed starch content. However, the importance of starch cannot be ruled out where the level of protein is the same in different genotypes. In addition to reserve food the strong correlation among seed

tissues show that embryo, endosperm and seed size are genetically interlinked traits that correlate with seed quality and seedling vigour traits. These findings are of great importance for the isolation of the corresponding genes and elucidation of the underlying mechanisms.

Introduction

Seed quality is defined as the viability and vigour attribute of a seed that enables the emergence and establishment of normal seedlings under a wide range of environments. Seeds are the major means of regeneration of plants and the suppliers of staple food of the human diet and our domesticated animals through the storage reserve of protein, starch and oil, synthesized during development and maturation (Bewley et al., 2012). The development of orthodox seeds proceeds through histo-differentiation and seed filling and terminates with a desiccation phase after which the embryo enters into a quiescent state, thereby permitting its storage and survival in various environmental conditions (Bewley et al., 2012).

The reserves are stored in different tissues of the seeds. In dicots most of the reserves are located within the embryo storage tissue including radicle, hypocotyl and the cotyledons while in monocots, including cereals, most of the storage reserves are accumulated in the endosperm. Dicots such as legumes generally store higher levels of protein (21-40%) and oil as compared to starch which is located mainly in the cotyledons. Contrary to dicots most monocot seeds contain higher levels of starch, located mainly in the endosperm and low levels of both protein and oil (Bewley et al., 2012). Tomato seed contains high levels (22-33%) of protein and lipid (20-29%) (Sheoran et al., 2005) and lower levels of starch (Schauer et al., 2005).

The storage compounds found in most mature seeds accumulate during the seed filling phase. They are principally storage proteins (mainly albumins, globulins, and prolamins), oil (often triacylglycerols) and carbohydrates (often starch; Baud et al., 2002). These reserves are of major importance as they support early seedling growth when metabolized upon germination and, therefore, participate in crop establishment. The success of establishment and vigour of the young seedlings is determined by the quality of the seed and its interaction with the environment. The food reserves it contains are available to sustain the seedling until it becomes an independent, autotrophic organism, able to use light energy.

Sucrose and amino acids are the major sources of carbon and nitrogen for the reserve synthesis of starch and protein in the seed. Both are imported in the reserve tissues (embryo and endosperm) from the photosynthetic organs of the parent plant. The quality and yield of the storage reserves is considerably influenced by the prevailing environmental conditions and the availability of carbon and nitrogen to the parent plant before and during

their synthesis. There appears to be an antagonistic relationship between starch and protein accumulation which advocates competition for carbon and nitrogen between the biosynthetic pathways of starch and storage proteins. An alternative explanation is that the carbon to nitrogen ratio functions as a signal to regulate these two biosynthetic pathways. In maize three kinds of relationships (dependent, independent and inverse) have been reported between the accumulation of starch and protein in the endosperm of maize kernels (Singletary and Below, 1989). The accumulation of starch and protein content in the endosperm increases concomitantly with the initial increase in the concentration of N in the medium (dependent relationship). With further increase in the concentration of N in the medium, there is a further increase in the accumulation of protein in the endosperm but no further increase or a decrease in the starch content (independent relationship). When the N supply is large enough, there is a decrease in the starch accumulation but not in protein content, and, thus, an inverse dependency between starch and protein production is observed (inverse relationship). In a similar fashion, protein accumulation is independent of carbohydrate supply where the synthesis of starch is. In wheat grains starch and protein synthesis are independent of each other (Donovan and Lee, 1977; Barlow et al., 1983). In the case of tomato, seed size has been analyzed in several studies and QTLs for seed size have been identified and in some cases fine mapped or even cloned (Doganlar et al., 2002; Khan et al., 2012), but little is known about the composition and protein and starch contents of tomato seeds.

Seedling vigour is essential for the emergence and establishment of healthy and normal seedlings. Good and uniform seedling establishment is vital for sustainable and profitable crop production and is therefore thought to be the most critical stage of a developing crop. Seedling vigour mostly depends on seed vigour and seed quality (Finch-Savage, 1995; Khan et al., 2012). Seed quality depends, among other things, on the amount and composition of protein, starch and oil, which are frequently defined as complex traits and are functionally dependent on C-N balance, central metabolism and sink-source interaction during development on the mother plant (Wobus and Weber, 1999; Toubiana et al., 2012). It is generally believed that large seeds produce vigorous seedlings and that this vigour is due to a higher amount of reserves in the larger seeds. Seed weight is considered to be an important attribute for the successful establishment, survival and vigour of seedlings (Poorter and Rose, 2005). Several studies have reported significant correlations between seed mass and seedling biomass acquired during a specific period of time after germination in different species, including tomato (Roy et al., 1996; Bonfil, 1998; Kidson and Westoby, 2000; Richards and Lukacs, 2002; Bettey et al., 2008; Khan et al., 2012). Co-location of QTLs for seed weight and seedling vigour traits have been described for various species (Alonso-Blanco et al., 1999; Bettey et al., 2000; Cui et al., 2002; Kehui et al., 2002; Finch-Savage et al., 2010; Khan et al., 2012). Seed size is often positively correlated with the protein content of the seed and, in turn, seed protein content is

frequently positively correlated with seedling vigour (Lowe and Ries, 1973; Ries and Everson, 1973; Evans and Bhatt, 1977; Saxena et al., 1987; Panthee et al., 2005). This implies that large-sized seeds will have a higher relative and total amount of protein and will produce more vigorous seedlings. Both the level and quality of protein are also affected by genotype and its interaction with environment and the concentration of seed protein is strongly correlated with seed weight and seedling vigour in wheat (Ries and Everson, 1973). In the case of starch, most studies have reported no or negative correlation between seed starch content and seed and seedling mass while in some crops grain starch content was positively correlated with grain weight and seedling biomass (Lai and McKersie, 1994; Cui et al., 2002; Sulpice et al., 2009; Ruffel et al., 2010).

In addition to correlation, co-location of QTLs for seed and seedling biomass and seed reserve have been detected in several genetic studies (Cui et al., 2002; Groos et al., 2003; Burstin et al., 2007). It has been shown that *APETALA2* (*AP2*) genes have a role in determining seed mass through regulation of both embryo cell number and size (Ohto et al., 2005) and in post embryonic organ formation through activation of stem cells of the primary shoot meristem (Wurschum et al., 2006). Petunia is a member of the *Solanaceae* family and the Petunia *phAP2* gene, which is an ortholog of the Arabidopsis *AP2* gene, can complement an Arabidopsis *ap2* mutant (Maes et al., 2001). In Arabidopsis, seed weight increases were accompanied by an increase in both seed protein and oil content in *ap2* mutants of Arabidopsis, indicating a common genetic mechanism involved in regulation of seed size and biosynthesis of seed protein (Jofuku et al., 2005). Seed size and mass and the chemical composition of seed storage compounds can vary within a species, variety or genetic line, or even within an individual plant. These variations in seed quality within the same genetic background are most likely caused by environmental factors prevailing during the growth period of the maternal plants, or by the position of seeds on the mother plant, both of which greatly affect the physiological and developmental characteristics of individual seeds. Seed size and mass are also determined by the genotype and its interaction with the environment and many QTLs and genes have been identified that control seed size in different species (Doganlar et al., 2000).

The seeds of tomato are composed of an embryo, endosperm and seed coat. Each of these three tissues has genetically distinct characteristics. The embryo develops from the fertilized ovule and contains an equal representation of the maternal and paternal genes (2N) whereas the endosperm contains two doses of the maternal parent's genes and one dose of the paternal parent's genes (3N). The endosperm serves as a source of food for the embryo during development and germination. The seed coat develops from the integument of the ovule and, therefore, contains only maternal genes. It protects the embryo and endosperm. The genetic balance and interaction between the endosperm, embryo and maternal tissues is a basic requirement for normal seed development and remains one of the most complex and unresolved issue of seed development. Although

embryo and endosperm size are closely related seed traits, they differentially correlate with seed weight and seedling vigour related traits in different crops species (Zhang and Maun, 1993) and different levels of accumulation of storage reserves have been documented in these two tissues of the seed (Singletary and Below, 1989; Lai and McKersie, 1994). Despite previous studies on the relationship between embryo and endosperm size and their relation with seed and seedling quality in food crops (López-Castañeda et al., 1996; Richards and Lukacs, 2002), no information is available about the relationship between embryo and endosperm size and their role in seed and seedling quality related traits in tomato.

The genetic dissection of seed processes contributing to size and mass, through QTL analysis, and their association with seed and seedling phenotypes, will help in unravelling the signalling pathways involved and will provide means to predict and enhance seed quality. Natural variation for seed reserve related traits existing in a recombinant inbred line (RIL) populations is a valuable resource to reveal the complex genetic mechanisms involved in the acquisition of seed quality (Ligterink et al., 2012). We used a RIL population generated from a cross between *S. lycopersicum* (cv. Moneymaker) and *S. pimpinellifolium* (Voorrips et al., 2000) to study these traits. This study demonstrates the presence of high quantitative variability in this population with respect to seed protein and starch content and, due to this variability, we were able to identify 24 QTLs related to seed protein and starch contents. The results also revealed weak to strong correlations between starch and protein contents and different seed and seedling vigour traits and co-location of QTLs for these traits. Comparing the 10 lines with the smallest seeds and the 10 lines with the biggest seeds also indicated significantly higher differences in the size of embryo and endosperm as well as in the tissue specific reserve contents of these seed tissues as compared to the differences in the whole RIL population. Further, the results also revealed strong correlation between embryo and endosperm masses and seed and seedling biomass as well as protein content of these tissues.

Materials and Methods

Plant material

The tomato RIL population was obtained from a cross between *Solanum lycopersicum* cv. Moneymaker and *Solanum pimpinellifolium* CGN 15528 (Voorrips et al., 2000). This population was genotyped for a total of 865 Single Nucleotide Polymorphism (SNP) markers in F7 and 100 RILs were produced in F8. 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 (AW van Heusden, personal communication) (Kazmi et al., 2012; Khan et al., 2012). The marker data were used to construct a genetic

linkage map consisting of 12 individual linkage groups corresponding to the 12 chromosomes of tomato.

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 population was grown under long day conditions (16h light and 8h dark) and the day and night temperatures were maintained at 25 and 15°C, respectively. The basic dose of fertilizer was uniformly supplied to all the RILs. 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 ($Na_3PO4.12H_2O$). 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). The cleaned seeds were dried for 3d 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 per plant and were used in the subsequent experiments.

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 of 2 X 100 seeds using a Nikon D80 camera with a 60 mm 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 analysed using the open source image analysis suite ImageJ (http://rsbweb.nih.gov/ij/) by using colour thresholds combined with particle analysis that automatically scored seed size (SS) as the area of selection in square pixels, circularity (SC) as $4p*(area/perimeter^2)$ and seed length (SL) as the longest distance between any two points along the selection boundary (feret's diameter). Seed size and seed lengths were also determined in 12-h imbibed seeds (ImbSS and ImbSL, respectively) (Khan et al., 2012).

Quantification of whole seed soluble protein and starch content

Protein extraction

We used 50 mg of dry seed per sample. The seed along with two bullets were put in round-bottom Eppendorf tubes and frozen in liquid nitrogen before grinding. After freezing each sample was ground for 4 minutes to a fine powder with the help of a Mikro-Dismembrator

U (B. Braun Biotech International). Subsequently, protein was extracted by a method as described by (Lohaus and Moellers, 2000) with some modifications. After grinding, 500 µl of the extraction buffer was added to each sample and vortexed till the powder was completely homogenized. Protein was extracted on ice for two hours. The extraction buffer consisted of 50mM HEPES pH 8.0, 1mM EDTA, 0.5M NaCl, 1 tablet of protease cocktail (Roche), 40 µl DNase(10U/µl) and 10µl RNase(10U/µl) per 10 ml extraction buffer. After extraction the samples were centrifuged for 20 minutes at 4°C at 14000 rpm. The clear supernatant was collected and transferred to new sterile Eppendorf tubes and stored on ice until further analysis.

Protein quantification

The concentrated supernatant was diluted 8 times with Milli-Q water. Further analysis was performed following the BIO RAD DC Protein assay instruction manual which is based on the Bradford method of protein analysis (http://wolfson.huji.ac.il/purification/PDF/Protein_Quantification/BIORAD_DC_Instr_Protein_Assay.pdf). The final absorbance value was measured in flat-bottom 96-well micro-plates (Greiner bio-one) with a SpectraMax Plus384 Absorbance Microplate Reader (Molecular Devices) at 750 nm. The protein content of the sample was quantified by the equation obtained from the standard curve by measuring known amounts of protein (BSA) with each measurement. The measurement was carried out in triplicate for three biological replicates for both harvests of the RIL population (2*3*3).

Starch extraction

For starch extraction we followed the same procedure for grinding the seeds as described for protein extraction and, after grinding, carbohydrates were extracted in 1 ml 80% methanol (MeOH) at 76°C for 15 minutes in water. After extraction the MeOH was evaporated in a Speedvac for two hours as described by Bentsink et al. (2000). The pellets left after extraction were then washed 4 times with Milli-Q water with the last wash carried out overnight to ensure complete removal of MeOH. After washing, the pellets were freeze dried in a freeze drier. Starch was extracted from the pellets left after carbohydrate extractions after extensive washing with water. Of the pellets (freeze dried powder) 20 mg was used for extraction of starch. The starch was extracted by adding 200 µl DMSO and 50 µl 8N HCl followed by shaking for 1h at 60°C in a water bath, then adding 300 µl Milli-Q water, 80 µl 8N NaOH and 370 µl citric acid buffer pH 4.5 followed by centrifuging for 1 minute at 14000 rpm. The clear supernatant was collected and transferred to new sterilized Eppendorf tubes. This supernatant was treated with 20 µl amyloglucosidase (AGS 1mg/mL) per 100 µl of sample and left overnight at 40°C to convert the starch into glucose. For the blank 20 µl of Milli-Q water was used instead of AGS.

Quantification of starch

Starch was determined as glucose. The glucose was measured using the HPLC-method for carbohydrate analysis on a Dionex DX500 HPLC-system using a Carbopac PA1-column + guard column at room temperature and detection by an ED40 electrochemical detector. A gradient of 20-150 mM NaOH was applied over 30 min (+10 min of 150 mM NaOH, if required). Between runs the column was washed for 5 minutes with 500 mM NaAc, followed by 10 minutes equilibration with 20 mM NaOH. The relative amount of starch (mg/g) was quantified with a standard curve obtained from measuring known amounts of starch. Measurements were carried out in triplicate for three biological replicates for both harvests of the RIL population (2*3*3).

Measurement of embryo and endosperm size, protein and starch content

To assess embryo and endosperm size and protein and starch content, we selected 20 extreme lines (10 lines with small and 10 with large seeds) of the RIL population on the basis of seed weight, including the two parents. The seeds were imbibed in water for 2 h and whole embryos were excised from the endosperm with the help of a scalpel. The embryo size was measured by taking close-up photographs of 200-400 embryos using a Nikon D80 camera as described above for measuring the seed size. The samples were stored at -20^{0}C. Embryo and endosperm weight was taken after freeze drying the samples in freeze drier. The protein and starch content of embryo and endosperm were quantified as described for whole seed protein and starch content. The analysis was carried out for each harvest in biological duplicates for embryo (2*2*3) and in triplicates for endosperm (2*3*3).

QTL mapping and statistical analysis

The QTL analysis was carried out with the help of a marker linkage map developed for the *S. lycopersicum* X *S. pimpinellifolium* RIL population as described above (Khan et al., 2012). The software package MapQTL®6.0 (Van Ooijen and Maliepaard, 2003) was used to identify and locate putative QTL positions in the genome for a given trait. A two-stage QTL analysis was performed. In the first stage, conventional interval mapping (Lander and Botstein, 1989) was performed at a 2 cM interval to detect putative QTLs. Secondly a multiple QTL mapping model (MQM) was used to identify the real QTLs (Jansen et al., 1995) as applied in MapQTL®6.0. In this method, background markers are selected as cofactors to reduce the residual variance. The logarithm-of-odds (LOD) profiles from interval mapping were inspected and the marker closest to each LOD peak of the putative QTLs (LOD ≥ 2.0) was selected as the cofactor to perform further MQM mapping analysis. The selection of cofactors resulted in new peaks in the LOD profiles, for which further cofactors were included in the analysis. Repeated rounds were performed to obtain the potentially

maximum number of cofactors for the MQM analysis. These cofactor markers were then subjected to backward elimination, as implemented in MapQTL®6.0, in order to select the best model for the second-stage MQM analysis. Such a backward elimination procedure leaves out one cofactor at a time in order to create a subset of cofactors (Yin et al., 2005). Each of these subset models is compared with the full model with all cofactors, and the subset model that causes the smallest change in likelihood is chosen as the starting set for a succeeding cycle of elimination. This process continued until the change in likelihood was significant according to a 0.002 P-value for the test. The set of cofactors then retained was used in the second stage of the MQM analysis. LOD threshold values applied to declare the presence of QTLs were estimated by performing the permutation tests implemented in MapQTL®6.0 using 1000 permutations of the original data set, resulting in a 95% LOD threshold of 2.0. The 1-LOD support intervals were established as QTL confidence interval (Van Ooijen, 1999) using the restricted MQM mapping procedure as implemented in MapQTL®6.0. The estimated additive genetic effects, percentage of variance explained by each QTL and the total variance explained by all the QTLs affecting a trait were obtained using MQM mapping.

Analysis of Heritability

The 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).

Data analysis

Pearson correlations between different traits were calculated with SPSS version 19.0. To generate the correlation matrix, the Pearson correlation analysis for all pairs of measured traits acros the whole population was performed by an R script (R version 2.13.1). Significant differences among all means of the RIL population as well as the extreme lines were estimated by General Linear Model, using Univariate analysis followed by a least significant difference (LSD) test. To compare embryo and endosperm size and protein and starch content of embryo and endosperm of the two groups of the extreme lines of the RIL population, the group analysis was performed by independent samples T-Test using SPSS version 19.0.

Results

Protein and starch content of whole seeds

To elucidate the underlying genetic basis of seed reserve food and its relation with seed and seedling vigour, we investigated the phenotypic variation in seed protein and starch contents of the 100 F_8 RILs derived from a cross between *S. lycopersicum* (*cv Moneymaker*) and *S. pimpinellifolium* (G1.1554). Seeds of the breeding line *S. lycopersicum* contained significantly (p-value <0.001) higher relative amounts of protein (mg/g seed, further referred to as RAP), as well as total amount of protein (µg/seed further referred to as TPS) than the wild accession *S. pimpinellifolium.* Both RAP ant TPS displayed significant (p-value <0.001) variation in whole seeds of the RIL population, mostly intermediary between the two parent lines, indicating segregation of high-protein loci from *S. lycopersicum* in the RIL population. In case of relative amount of starch (mg/g of seed, further referred to as RAS) we observed no significant difference between the two parents. As a result, the total amount of starch per seed (µg/seed, further referred to as TSS) of *S. lycopersicum* was significantly higher than the *S. pimpinellifolium parent* (p-value<0.001). On the other hand we observed highly significant (p-value <0.001) variation in both RAS and TSS in the RIL population. In addition, a strong transgressive segregation was observed for the RAS, as nearly half of the RIL population had significantly higher and about 22% of the RILs had significantly lower levels of RAS compared to the parents (Table 4.1, Figure 4.1). Estimation of heritability indicated that the RAP, TPS, RAS and TSS of the RIL population are highly heritable (Table 4.2). The RIL population showed normal distribution for RAP and strong positive skew for TPS, RAS and TSS (Figure 4.1)

Table: 4.1. Phenotypic and quantitative analysis of seed reserve food of the *S. lycopersicum* x *S. pimpinellifolium* RIL population, including the two parents.

S. No	Trait[1]	*S. lycopersicum*[2]	*S. pimpinellifolium*[3]	RIL Population[4]	F-Value[5]	P-value[6]
		Mean±SD	Mean±SD	Mean±SD		
1.	RAP	99.23±2.79	71.76±2.38	86.51±10.15	7.936	<0.001
2.	TPS	291.04±8.18	77.91±2.58	148.66±41.2	51.818	<0.001
3.	RAS	5.16±0.76	5.69±0.76	6.85± 3.99	3.744	<0.001
4.	TSS	15.14±2.40	6.18±0.82	11.75±7.67	4.050	<0.001

1. Traits, RAP= Relative Amount of Protein (mg/g of seed), TPS= Total amount of Protein (µg/seed), RAS= Relative Amount of Starch (mg/g of seed), TSS= Total amount of Starch (µg/seed).
2. Mean and standard deviation values of *S. lycopersicum* parent.
3. Mean and standard deviation values of *S. pimpinellifolium* parent
4. Mean and standard deviation values of the RIL population.
5. F values for one way ANOVA.
6. P values for one way ANOVA.

Correlations between reserve food and seed and seedling phenotypes

To explore the relationship between the reserve food of whole seed measured in the present experiment and the seed and seedling phenotypes measured during our previous study (Khan et al., 2012), a correlation matrix was created by executing Pearson correlation analysis for all pairs of measured traits across the whole population. In this analysis we used the average values of raw data calculated for a given trait/RIL pair. The Pearson correlation coefficients (Rp) and accompanying false discovery rate (FDR)-corrected *P*-values (P_{BH}; (Benjamini and Yekutieli, 2001) are provided in Supplemental Table S4.1. Using the Pearson correlation coefficient to calculate associations among amounts of reserve food and seed or seedling quality phenotypes, a number of low to high significant correlations were observed between RAP of whole seed and seed and seedling quality traits (Figure 4.2; Supplemental Table S4.1). This correlation was most obvious ($P_{BH} < 0.02$) with MRL (Main Root Path Length), TRS (Total Root Size), SW (Seed Weight), ImbSS (Imbibed Seed Size and the seedling vigour traits of seedlings grown under nutrient stress conditions FrRtWn (Fresh Root weight Without nutrients), DrRtWn (Dry Root weight Without nutrients), FrShWn (Fresh Shoot weight Without nutrients), and DrShWn (Dry Shoot weight Without nutrients). These correlations were even stronger between TPS and the phenotypic traits of seed and seedling. In this case the correlation was also significant with traits of seedlings grown under normal nutrient conditions FrRt (Fresh Root weight), DrRt (Dry Root weight) , FrSh (Fresh Shoot weight) and DrSh (Dry Shoot weight) ($P_{BH} < 0.0001$). In the case of starch content of whole seed, we could not find any correlation between RAS and seed and seedling phenotypes. On the other hand, highly significant correlation was observed between TSS and most seed and seedling vigour traits ($P_{HB} < 0.001$). This reveals that although there is no link between the concentration of starch and seed weight, the large seeds have higher amount of total starch per seed as compared to small seeds. Although there is a significant correlation between SW and root system architecture (RSA) parameters such as main root path length (MRL), total root size (TRS), lateral root number (LRn) and lateral root density per basal zone (LRD_Bz), we could not find any significant correlation between TSS and RSA traits (Figure 4.2, Supplemental Table S4.1).

Identification of QTLs for protein and starch reserves

We used the normalized data of whole seed reserves to map QTLs. The map position and characteristics of the QTLs associated with the studied seed reserve phenotypes are summarized in Table 4.2. Permutation tests revealed a LOD threshold of 2.0 and this threshold was used to determine putative QTLs where seed reserve phenotypes map. By generating a heat map of LOD profiles, QTLs can be visualized and global 'hot spots' and empty regions across the 12 chromosomes can be seen (Figure 4.3). Multiple QTL (MQM) mapping analysis revealed a total of 24 significant QTLs on 15 different positions for the 4

seed reserve food traits tested across the RIL population (Table 4.2). Two QTLs were identified for the relative amount of protein (RAP), with the major QTL on chromosome 9 and a minor QTL on chromosome 4 with a total explained variance of 29.2%. Six QTLs, one each on chromosomes 1, 2, 4, 7, 9 and 11 were detected for the total amount of protein (TPS) with a total explained variance of 53.1%. For all the QTLs related to TPS as well as the major QTL for RAP, the favourable alleles were derived from the *S. lycopersicum* parent (negative additive effects in Table 4.2 and yellow to red colour in the heat map of Figure 4.3), revealing a link of protein content with large seed size. In case of the relative amount of starch (RAS) we identified a higher number of QTLs (7 QTLs) as compared to the number of QTLs detected for RAP (2 QTLs), thus revealing the complexity of the RAS is regulated by complex additive effect of many QTLs.

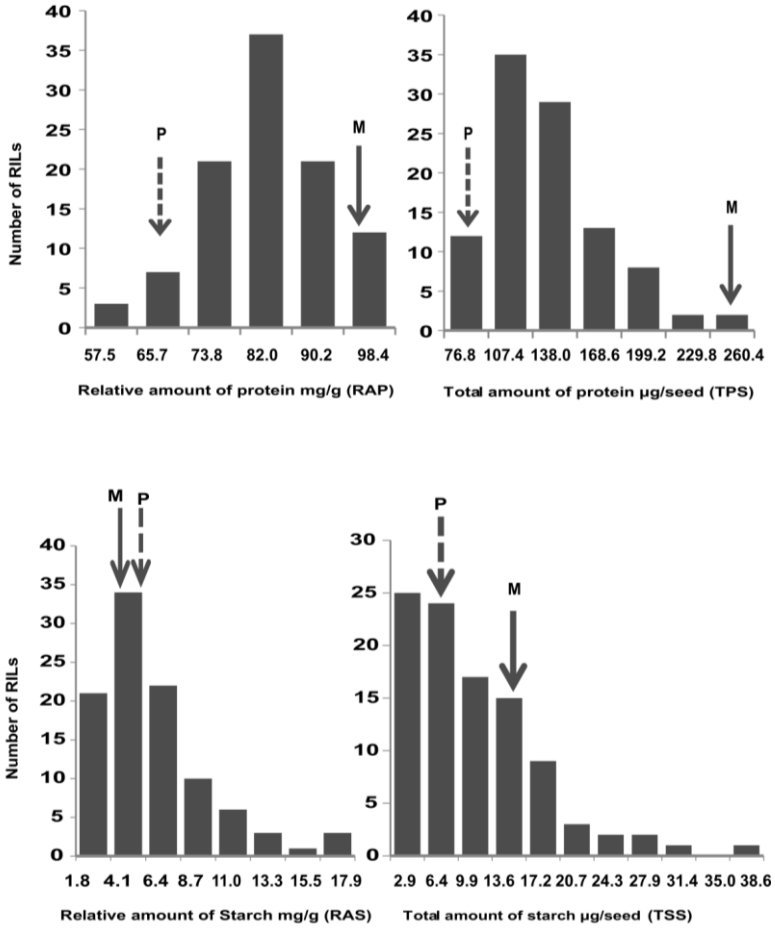


Figure 4.1. Frequency distributions of non-normalized data of relative amount (mg/g) and total amount (µg/seed) of protein and starch content of whole seed phenotypes in the *Solanum lycopersicum* and *Solanum pimpinellifolium* recombinant inbred line (RIL) population. The average parental value is indicated with a solid arrow (M) and dotted arrow (P). P= *S. pimpinellifolium* parent and M= *S. lycopersicum* parent.

Table 4.2. Overview of significant QTLs associated with seed and seedling traits of the *S. lycopersicum* x *S. pimpinellifolium* RIL population.

Traits[1]	Chr[2]	Confidence Interval (cM)[3]	Nearest Marker[4]	LOD Score[5]	Additive Effect[6]	Explained Variance (%)[7]	Total Variance (%)[8]	Heritability[9]
RAP	4	23.15-43.36	3143736	2.33	0.589	7.8	29.2	0.89
	9	59.94-86.20	61607962	5.82	-0.985	21.4		
TPS	1	75.31-101.75	72402911	5.44	-0.846	14.0	53.1	0.98
	2	55.89-94.81	43635207	2.41	-0.515	5.7		
	4	41.10-88.07	53711645	2.43	-0.501	5.8		
	7	54.39-74.72	61450244	2.55	-0.524	6.1		
	9	65.00-79.01	62098389	6.09	-0.847	16.0		
	11	0.00-21.38	5148492	2.30	-0.490	5.5		
RAS	1	8.32-119.88	72402911	2.56	-0.585	4.9	53.8	0.79
	2	16.26-26.99	33753248	5.79	-0.874	12.2		
	2	54.31-66.89	42037927	5.55	0.723	11.3		
	4	0.00-16.19	516524	3.49	0.566	6.9		
	7	34.46-51.12	55492731	3.43	-0.543	6.7		
	11	53.34-72.23	52219825	2.65	-0.492	5.1		
	12	27.21-49.64	3920250	3.47	0.579	6.7		
TSS	1	2.03-19.62	1226398	2.88	-0.351	5.0	66.5	0.80
	1	73.31-101.75	72402911	7.86	-0.776	12.2		
	2	12.26-27.99	34231799	2.93	-0.786	7.5		
	2	51.33-66.89	41701555	2.35	0.565	6.6		
	4	0.00-8.19	516524	5.49	0.738	11.1		
	6	15.97-29.48	31371391	3.20	0.500	4.4		
	7	37.46-45.47	55492731	7.68	-0.751	12.0		
	8	99.69-109.84	62657015	3.40	0.494	4.7		
	12	18.52-52.32	3471392	2.24	0.393	3.0		

[1]Traits, RAP= Relative amount of protein (mg/g seed), TPS= total amount of protein (µg/seed), RAS = Relative amount of starch (mg/g seed) and TSS = Total amount of starch (µg/seed). [2]Chromosomes on which the QTLs were detected. [3]1-LOD support interval in centi-Morgan. [4]Nearest marker to the position of ìdentified QTL. [5]LOD score (LOD score of 2 or above was calculated to be significant for this population). [6]Additive effect; 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. [7]Percentage of variation explained by each QTL. [8]Percentage of total variation explained by genetic factors for a single trait as estimated by MapQTL. [9]Broad-sense heritability estimate for each trait.

These QTLs had a total explained variance of 53.8% and were located one each on chromosomes 1, 4, 7, 11 and 12 and two on chromosome 2. Nine QTLs were related to the total amount of starch (TSS) with a total explained variance of 66.5%. These QTLs were mostly the same as identified for the RAS, except three QTLs, one each positioned on chromosomes 1, 6 and 8. For nearly half of the QTLs related to RAS and TSS the favourable alleles were derived from *S. lycopersicum* and the other half from *S. pimpinellifolium* (negative and positive additive effects in Table 4.2), demonstrating the control of starch by loci other than the ones that control seed weight where all the favourable alleles were derived from *S. lycopersicum*.

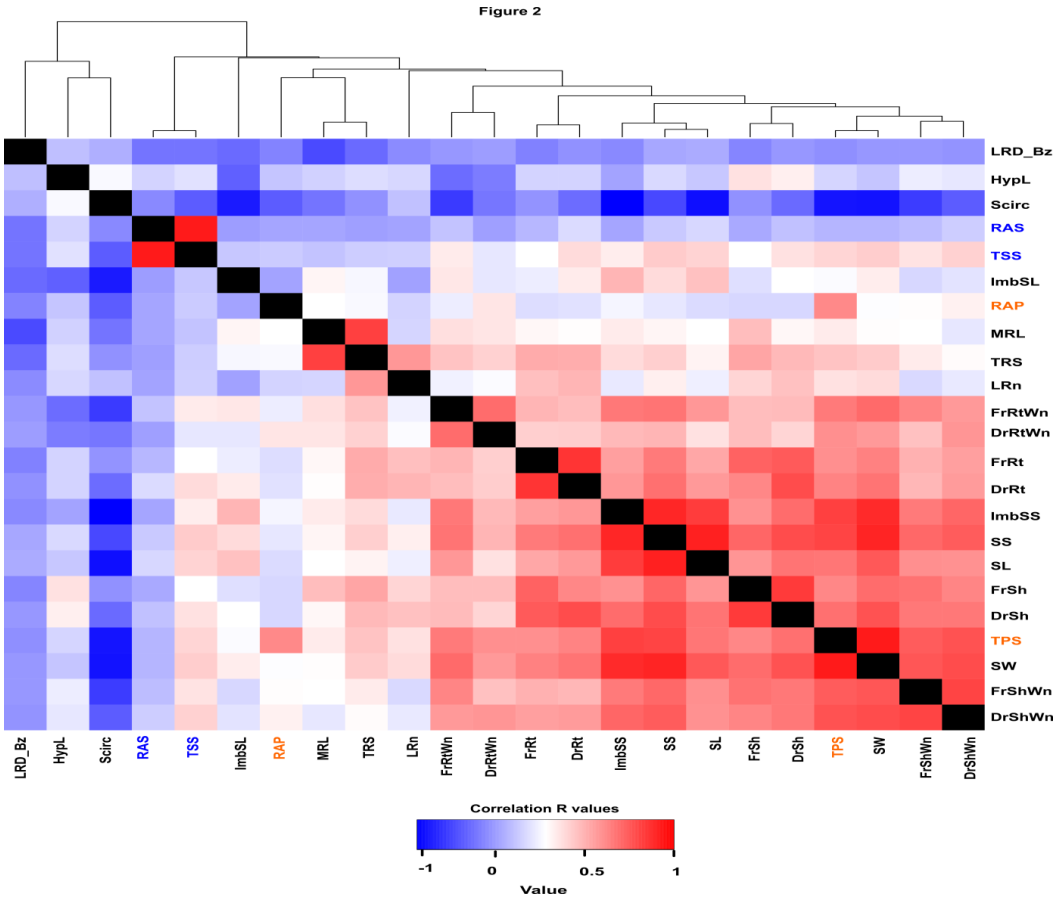


Figure 4.2. Pearson correlation of whole seed protein and starch content with seed, seedling and RSA traits. RAP= relative amount of protein (mg/gram seed), TPS= total amount of protein (µg/seed), RAS= relative amount of starch (mg/gram seed) and TSS= total amount of starch (µg/seed). SW= Seed Weight, SS= Seed Size, SL= Seed Length, FrRt= Fresh Root weight, DrRt= Dry Root weight, FrSh= Fresh Shoot weight, and DrSh= Dry shoot weight, FrShWn= Fresh Shoot weight without nutrients, DrShwn= Dry shoot weight without nutrients, FrRtWn= Fresh Root weight without nutrients, DrRtWn= Dry Root weight without nutrients.

Identification of QTLs for protein and starch reserves

We used the normalized data of whole seed reserves to map QTLs. The map position and characteristics of the QTLs associated with the studied seed reserve phenotypes are summarized in Table 4.2. Permutation tests revealed a LOD threshold of 2.0 and this threshold was used to determine putative QTLs where seed reserve phenotypes map. By generating a heat map of LOD profiles, QTLs can be visualized and global 'hot spots' and empty regions across the 12 chromosomes can be seen (Figure 4.3). Multiple QTL (MQM) mapping analysis revealed a total of 24 significant QTLs on 15 different positions for the 4 seed reserve food traits tested across the RIL population (Table 4.2). Two QTLs were identified for the relative amount of protein (RAP), with the major QTL on chromosome 9 and a minor QTL on chromosome 4 with a total explained variance of 29.2%. Six QTLs, one each on chromosomes 1, 2, 4, 7, 9 and 11 were detected for the total amount of protein (TPS) with a total explained variance of 53.1%. For all the QTLs related to TPS as well as the major QTL for RAP, the favourable alleles were derived from the *S. lycopersicum* parent (negative additive effects in Table 4.2 and yellow to red colour in the heat map of Figure 4.3), revealing a link of protein content with large seed size. In case of the relative amount of starch (RAS) we identified a higher number of QTLs (7 QTLs) as compared to the number of QTLs detected for RAP (2 QTLs), thus revealing the complexity of the RAS is regulated by complex additive effect of many QTLs. These QTLs had a total explained variance of 53.8% and were located one each on chromosomes 1, 4, 7, 11 and 12 and two on chromosome 2. Nine QTLs were related to the total amount of starch (TSS) with a total explained variance of 66.5%. These QTLs were mostly the same as identified for the RAS, except three QTLs, one each positioned on chromosomes 1, 6 and 8. For nearly half of the QTLs related to RAS and TSS the favourable alleles were derived from *S. lycopersicum* and the other half from *S. pimpinellifolium* (negative and positive additive effects in Table 4.2), demonstrating the control of starch by loci other than the ones that control seed weight where all the favourable alleles were derived from *S. lycopersicum*.

Traits were regarded as co-locating when the 1-LOD confidence intervals of the traits overlapped with each other. Moreover, traits were assumed pleiotropic when the direction of the effect of the QTLs for the co-locating traits was the same (negative or positive additive effect, Table 4.2, and yellow to red or light blue to dark blue colours in the heat map of Figure 4.3). In our previous study (Khan et al., 2012) we observed that the QTLs for seed traits were co-locating with QTLs for seedling vigour traits. To investigate whether the QTLs of reserve food also co-locate with seed and seedling trait QTLs, we compared the QTL profiles of reserve food with the QTL profiles of seed and seedling traits (Figure 4.3 and Supplemental Table S4.2). The results revealed that most of the QTLs for RAP and TPS were co-locating with seed and seedling vigour trait QTLs. For example, the major QTL for RAP on chromosome 9 (RAP_9) co-locates with 14 seed and seedling trait QTLs (FrRt, DrRt, FrSh DrSh, SW, SS, ImbSS, ImbSL (Imbibed seed length), FrRtWn, DrRtWn,

FrShWn, DrShWn, MRL and TRS), whereas the minor QTL RAP_4 does not co-locate with any seed or seedling trait QTL. Strikingly, all six TPS QTLs are co-locating with seed and seedling trait QTLs. These include TPS_1, co-locating with 12 seed and seedling trait QTLs on chromosome 1 (FrRt, DrRt, FrSh DrSh, SW, SS, SL, ImbSS, FrRtWn, DrRtWn, FrShWn, and DrShWn) and also with the RAS_1 and TSS_1 QTLs. TPS_2 is co-locating with SW, ImbSL, FrShWn, DrShWn and LRD_Bz QTLs on chromosome 2. TPS_4 is co-locating with 8 seed and seedling trait QTLs on chromosome 4 (DrRt, DrSh, SW, SS, SL, ImbSS, ImbSL, and DrShWn). The QTL on chromosome 7 (TPS_7) is co-locating with 6 seed and seedling trait QTLs (SW, SS, SL, DrRtWn, FrShWn, DrShWn). TPS_9 is co-locating with the same 14 seed and seedling trait QTLs as described for RAP_9 above and TPS_{11} is co-locating with 4 QTLs on chromosome 11 (SW, SL, ImbSS, and TRS) (Table 4.2, Figure 4.3, and Supplemental Table S4.2).

In the case of seed starch content, there are three QTLs that co-locate both with seed and seedling traits and eight QTLs that co-locate only with seedling traits, in addition to 5 QTLs that do not co-locate with any seed or seedling trait. For example, the QTLs RAS_1 and TSS_1 co-locate with the same 12 seed and seedling trait QTLs on chromosome 1 as revealed for TPS_1 above. Similarly, TSS_7 also co-locates with both seed and seedling trait QTLs on chromosome 7, as described for TPS_7 above.

Co-location of QTLs for food reserves and seed quality and seedling vigour

The examples of QTLs that only co-locate with seedling traits include $RAS_{2\text{-}1}$, $RAS_{2\text{-}2}$, $TSS_{2\text{-}1}$ and $TSS_{2\text{-}2}$ on chromosome 2, which co-locate with the QTLs for LRn (Lateral Root number), MRL and TRS. Likewise, RAS_4, TSS_4 and TSS_6 co-locate with FrSh and LRn on chromosome 4 and 6 respectively, whereas RAS_7 overlaps with the QTLs for DrRtWn and FrShWn on chromosome 7. The third class of QTLs that do not co-locate with any seed or seedling trait include RAS_6, TSS_8, RAS_{11}, RAS_{12} and TSS_{12} on chromosomes 6, 8,11 and 12 respectively (Figure 4.3, Table 4.2 and Supplemental Table S4.2).

Embryo and endosperm traits

To find out whether seed size and/or weight depend on embryo or endosperm characteristics, or both, and how the reserves are distributed over the different seed tissues, we separated embryo from endosperm (+ testa). Embryo weight (in mg per seed, further referred to as Emb_Wt) accounted for 30-40% of the seed weight, whereas endosperm + testa (in mg per seed, further referred to as End_Wt) accounted for 60-70% of the seed weight. Univariate analysis revealed significant differences among the means of Emb_Wt and End_Wt, and embryo size (Emb_S) of the extreme lines. As expected, an independent sample t-test to compare the two groups demonstrated that the group of extreme lines with high seed weight had significantly (p value <0.001) greater Emb_Wt,

End_Wt and Emb_S than the group with low seed weight (Table 4.3, Figure 4.4). Although there was some dissimilarity, nearly all the data of Emb_Wt, Emb_S and End_Wt followed the same trends as SW when sorted by ascending order of SW of the extreme lines (Figure 4.4).

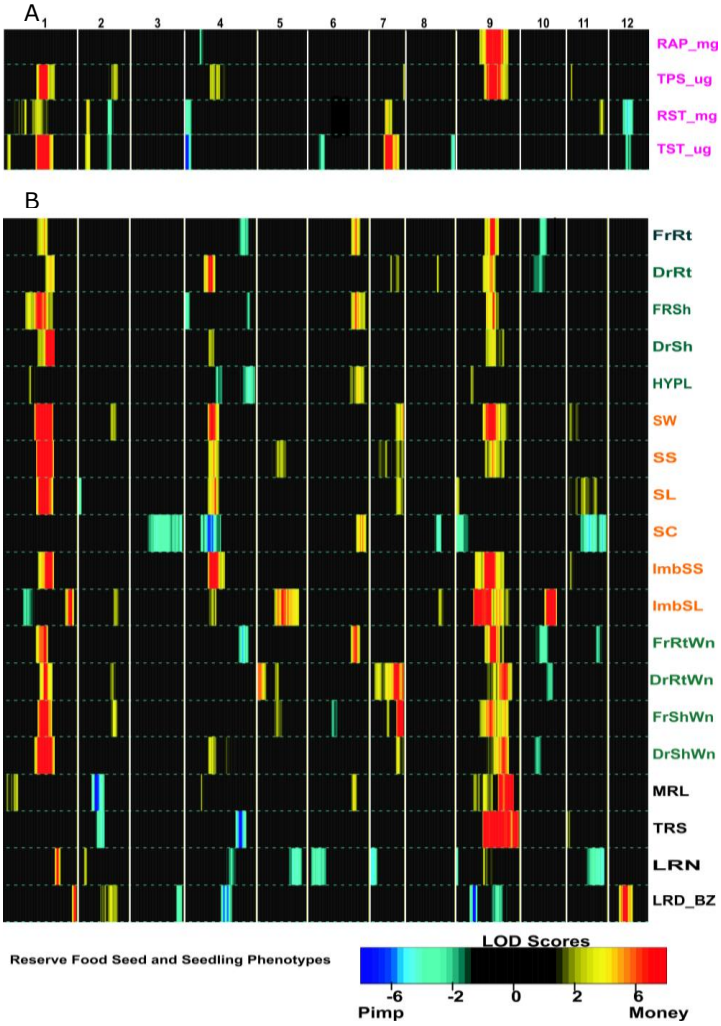


Figure 4.3. Heatmap of QTLs identified for seed reserve food phenotypes (**A**) and their co-location with seed and seedling quality traits QTLs previously identified by Khan, Kazmi, et al (2012) (**B**). Tomato chromosomes are identified by numbers (1–12), with centimorgans ascending from the left to right; chromosomes are separated by white lines. RAP= relative amount of whole seed protein (mg/gram seed), TPS= total amount of protein (µg/seed), RAS= relative amount of starch (mg/gram seed) and TSS= total amount of starch (µg/seed). SW= Seed Weight, SS= Seed Size, SL= Seed Length, FrRt= Fresh Root weight, DrRt= Dry Root weight, FrSh= Fresh Shoot weight, and DrSh= Dry shoot weight, FrShWn= Fresh Shoot weight without nutrients, DrShwn= Dry shoot weight without nutrients, FrRtWn= Fresh Root weight without nutrients, DrRtWn= Dry Root weight without nutrients, 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 allelic effect for the trait in *S. pimpinellifolium*, and yellow and red in *S. lycopersicum*.

Tissue-specific food reserves

To assess the distribution of reserves in embryo and endosperm, protein and starch contents were determined in both tissues. The data for the means, standard deviation (SD), F-values for the univariate and t-value for the independent sample t-test and p-values for the tissue-specific food reserves are presented in Table 4.3. Univariate analysis revealed significant (p values ≤ 0.001) differences among the means of the extreme lines for relative and total amounts of protein and starch of embryo and endosperm (RAP_Emb, RAP_End, TPS-Emb, TPS_End, RAS-Emb, RAS_End, TSS_Emb and TSS_End). The *S. lycopersicum* parent had significantly higher RAP and TPS as compared to the *S. pimpinellifolium* parent. Independent sample t-test indicated that both embryo and endosperm of the large seed group had significantly higher (p value < 0.001) RAP and TPS as compared to the embryo and endosperm of the small seed group (Table 4.3, Figure 4.4). Further, the RAP, as well as TPS of the embryo were significantly higher (p value <0.001) than the RAP and TPS of endosperm in both seed size groups. In the case of starch, the *S. pimpinellifolium* parent had significantly higher (p value <0.001) RAS_Emb than the *S. lycopersicum* parent. However there was no significant difference between the TSS_Emb of the two parents. The independent sample t-test revealed that the RAS_Emb of the small-size seed group were significantly higher (p value 0.03) than the RAS_Emb of the large group, but the two groups were not different with respect to TSS_Emb. Although there were highly significant differences among the RAS_End of the extreme lines, the group analysis revealed no significant difference between the two groups, indicating greater variation within the groups. Consequently, the TSS_End of the large-seed group was significantly higher (p value 0.001) than the TSS_End of the small-seed group. Similar trends were observed between the two parents where *S. lycopersicum* (large seed size) had significantly higher (p value 0.04) TSS_End than the *S. pimpinellifolium* parent (small-seed size).

Relationship of embryo, endosperm and their food reserves with seed and seedling phenotypes of the extreme lines

To investigate possible associations among embryo, endosperm and their relative and total amounts of reserve food and seed quality and seedling vigour traits measured during our previous studies (Kazmi et al., 2012; Khan et al., 2012), a correlation matrix was created by measuring Pearson correlations for all pairs of measured traits across the 20 extreme lines of the RIL population, including the parents. In this analysis we used the average values of raw data calculated for a given trait/RIL pair. The Pearson correlation coefficients (Rp) and accompanying false discovery rate (FDR)-corrected *P*-values (P_{BH}; (Benjamini and Yekutieli, 2001) are provided in Supplemental Table S4.3. To interpret and visualize the relationship among different traits, a correlation plot based on the Pearson correlation coefficient (Rp) is given in Figure 4.5. This plot shows strong significant correlations (Rp values 0.5 to 0.95)

Table 4.3. Phenotypic and reserve food analysis of embryo and endosperm parameters of the 20 extreme lines of the RIL population, including the two parents

4.3.A. One Way ANOVA among the means of the extreme lines

S. No	Traits[1]	*S. lycopersicum*	*S. pimpinellifolium*	Extreme Line	F-Value	P-value[5]
		Mean	Mean	Mean±SD		
1.	Emb_Wt	2.051±0.07	0.6±0.03	1.09±0.42	53.24	<0.001
2.	End_Wt	3.48±0.33	1.53±0.016	2.14±0.87	4.19	0.001
3.	Emb_S	4.75±0.11	2.69±0.02	3.22±0.75	71.52	<0.001
4.	RAP_Emb	166.68±3.28	113.03±0.51	131.76±24.10	19.31	<0.001
5.	RAP_End	79.38±8.69	68.19±3.57	74.57±12.72	4.12	0.001
6.	TPS_Emb	488.87±9.61	122.711± 0.55	230.62±113.99	122.46	<0.001
7.	TPS_End	232.82±25.49	74.04±3.88	126.79±54.21	28.91	<0.001
8.	RAS_Emb	1.59±0.28	4.69±0.01	2.82±2.09	21.07	<0.001
9.	RAS_End	1.23±0.35	1.73±0.06	2.05±0.82	6.80	<0.001
10.	TSS_Emb	4.67±0.31	5.09±0.03	4.46±3.04	11.34	<0.001
11.	TSS_End	3.59±1.74	1.88±0.38	3.41±1.74	8.77	<0.001

4.3.B. Independent Sample T-Test between the two groups

S. No	Traits[1]	Low 10%[2]	High 10%[3]	Extreme Line[4]	t-Value	P-value[5]
		Mean±SD	Mean±SD	Mean±SD		
1.	Emb_Wt	0.74±0.12	1.41±0.34	1.09±0.37	-8.215	<0.001
2.	End_Wt	1.36±0.19	2.84±0.61	2.10±0.805	-10.414	<0.001
3.	Emb_S	2.62±0.28	3.76±0.63	3.19±0.690	-7.416	<0.001
4.	RAP_Emb	109.92±8.16	151.62±14.37	130.78±22.96	-11.403	<0.001
5.	RAP_End	67.19±11.95	81.27±9.71	74.23±11.72	-4.497	<0.001
6.	TPS_Emb	127.72±24.15	324.16±74.79	225.94±101.1	-11.216	<0.001
7.	TPS_End	77.20±15.45	171.88±	124.54±54.81	-12.115	<0.001
8.	RAS_Emb	3.55±2.54	2.17±1.34	2.86±2.11	2.246	0.030
9.	RAS_End	2.13±0.95	1.98±0.70	2.05±0.771	0.585	0.562
10.	TSS_Emb	4.26±3.26	4.64±2.89	4.45±3.02	-0.141	0.889
11.	TSS_End	2.54±1.34	4.21±1.71	3.37±1.71	-3.494	0.001

[1]Traits, Emb_Wt= embryo weight (mg/seed), End_Wt =Endosperm weight (mg/seed), Emb_S= Embryo size (mm), RAP_Emb= Relative amount of embryo protein (mg/gram), RAP_End= Relative amount of endosperm protein (mg/gram), TPS_Emb= Total amount of embryo protein (µg/embryo), TPS_End= Total amount of endosperm protein (µg/endosperm), RAS_Emb= Relative amount of embryo starch (mg/gram), RAS_End= Relative amount of endosperm starch (mg/gram), TSS_Emb= Total amount of embryo starch (µg/embryo), TSS_End= Total amount of endosperm starch (µg/endosperm). [2]Top 10% of RILs in the population with the lowest seed weight. [3]Top 10% of RILs in the population with the highest seed weight. [4]Mean and standard deviation of the two groups combined. [5]P values for one way ANOVA, indicating significant difference in the extreme RILs including the two parents (Table 4.3.A.) and P values for independent sample t-test showing significant differences in the two groups of the RIL population (Table 4.3B).

of embryo weight per seed (Emb_Wt), endosperm weight per seed (End_Wt) and embryo size (Emb_S) with all seed and seedling phenotypes. However, the correlation of both Emb_Wt and Emb_S is stronger than the correlation of End_Wt with almost all the seed and seedling traits with the greatest difference in correlation with hypocotyl length (Supplemental Table S4.3).

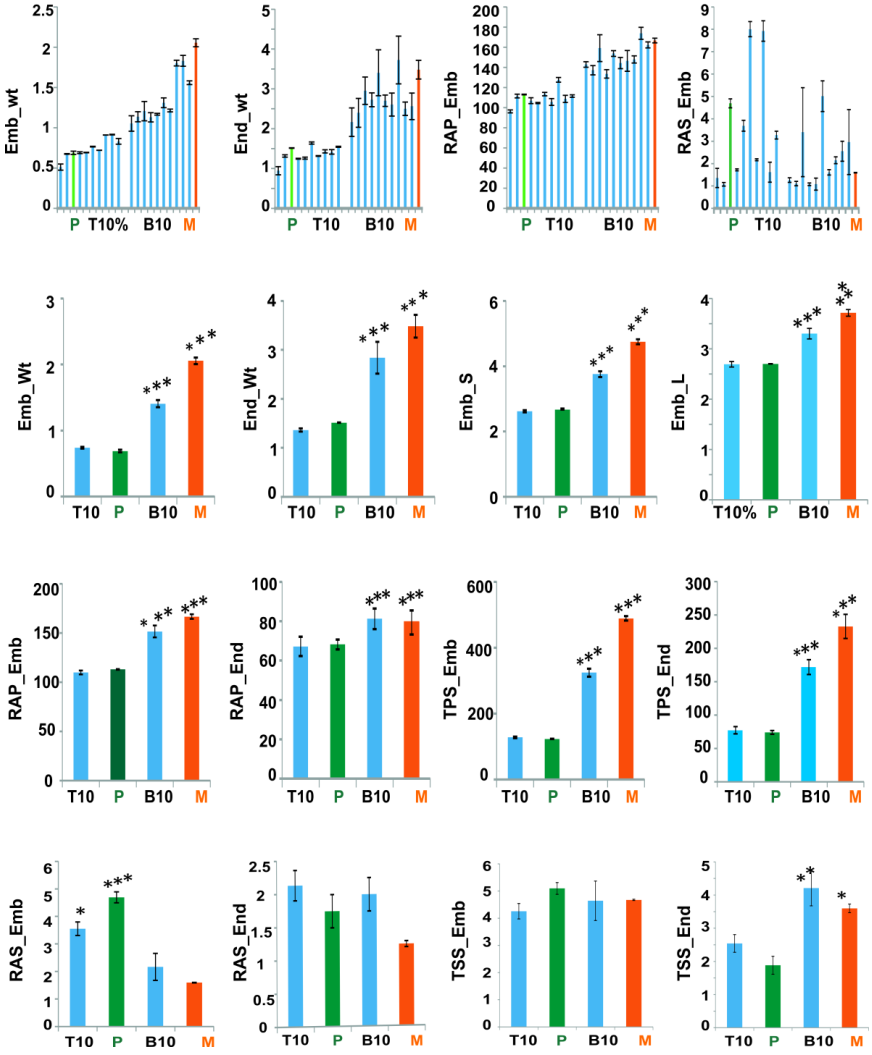


Figure 4.4. Seed tissue weight and tissue specific reserve food phenotypes. The data are the means + standard errors of two biological and 3 technical replicates for each harvest for embryo (2*2*3) and 3 biological and 3 technical replicates for each harvest for endosperm (2*3*3).

First row: Emb_wt= Embryo weight (mg/seed), End_wt= Endosperm weight (mg/seed), RAP_Emb= Relative Amount of Protein of embryo (mg/gram embryo), RAS_Emb= Relative Amount of Starch (mg/gram embryo), T10= top 10% RILs of the *S. lycopersicum* x *S. pimpinellifolium* RIL population, B10= Bottom 10% RILs, P= *S. pimpinellifolium* parent and M= *S. lycopersicum* cv. Moneymaker parent.

Second row: Emb_wt= Embryo weight (mg/seed), End_wt= Endosperm weight (mg/seed), Emb_S= average Embryo Size (mm), Emb_L= average Embryo Length (mm).

Third row: RAP_Emb= average Relative Amount Protein in Embryo (mg/gram of embryo), RAP_End= average Relative Amount of Protein in Endosperm (mg/gram of endosperm), TPS_Emb= average Total amount of Protein per embryo (µg/embryo), TPS_End= average Total amount of Protein per endosperm (µg/endosperm).

Fourth row: RAS_Emb= average Relative Amount of Starch in Embryo (mg/gram of embryo), RAS_End= average Relative Amount of Starch in Endosperm (mg/gram endosperm), TSS_Emb = average Total amount of Starch per Embryo (µg/embryo), TSS_End = average Total amount of Starch per Endosperm (µg/endosperm).

In addition, both the relative (RAP) and total amount of protein (TPS) of embryo and endosperm revealed similar trends where they had significantly higher correlation with most of the seed and seedling traits and the correlation of both RAP and TPS of embryo was stronger than the RAP and TPS of endosperm (Figure 4.5 and Supplemental Table S4.3). Moreover, the RAP of endosperm had no correlation with the hypocotyl length, whereas the correlation of the RAP of the embryo strongly correlated with hypocotyl length (B_{PH} 0.012 and p value 0.007). In case of tissue-specific starch of the extreme lines, the RAS of both embryo and endosperm did not indicate any association with seed and seedling traits. However, the TSS of the endosperm displayed significant correlation with most of the seed and seedling traits, including the embryo and endosperm phenotypes of the extreme lines (B_{PH} values ranging from 0.01 to 0.001). Conversely, the TSS of the embryo did not reveal any correlation with any of the seed and seedling traits. Furthermore, the protein and starch contents of the specific tissues had also significant correlation with the protein and starch content of the whole seed, displaying consistency across the different experiments.

Discussion

In our previous studies (Kazmi et al., 2012; Khan et al., 2012) we analyzed 83 lines of the *S. lycopersicum* x *S. pimpinellifolium* RIL population for seed germination and seed and seedling vigour related traits and identified 120 QTLs for seed germination traits and 62 QTLs for seed size and seedling vigour traits. In the seed size and seedling vigour study, we demonstrated significant correlation between seed dimensions, seedling biomass and root architecture and most of the QTLs for seed dimensions were co-locating with QTLs for seedling biomass under both normal and nutritional stress conditions. We hypothesized that seedling vigour was related to higher amounts of seed reserves in the larger seeds. If so, then the quantity of food reserves per single seed should correlate with the vigour and biomass of the related seedling and the QTLs for seed dimensions and seedling vigour should co-locate with QTLs for seed reserves. To test this hypothesis, in this study we analyzed 100 lines of the same RIL population by including 17 extra lines of this population which were not genotyped at the time of our previous studies. In agreement with our previous studies, we observed large variation between the two parents for all the phenotypic traits regarding seed reserves, except the relative amount of starch although there were some highly significant differences in the RIL population. However, these differences for the relative amount of starch in the RIL population were due to strong transgression observed in the RIL population, as half of the RIL population had significantly higher and about 22% had significantly lower RAS as compared to the two parents while the remaining 30% of the RILs were in the same range as the two parents (Figure 4.1). Compared to starch, no transgressive segregation was observed for the protein content of

seed. However, our hypothesis is supported, as we found significant correlation and co-location of QTLs among amounts of reserve food and seed quality and seedling vigour related traits which will be discussed below.

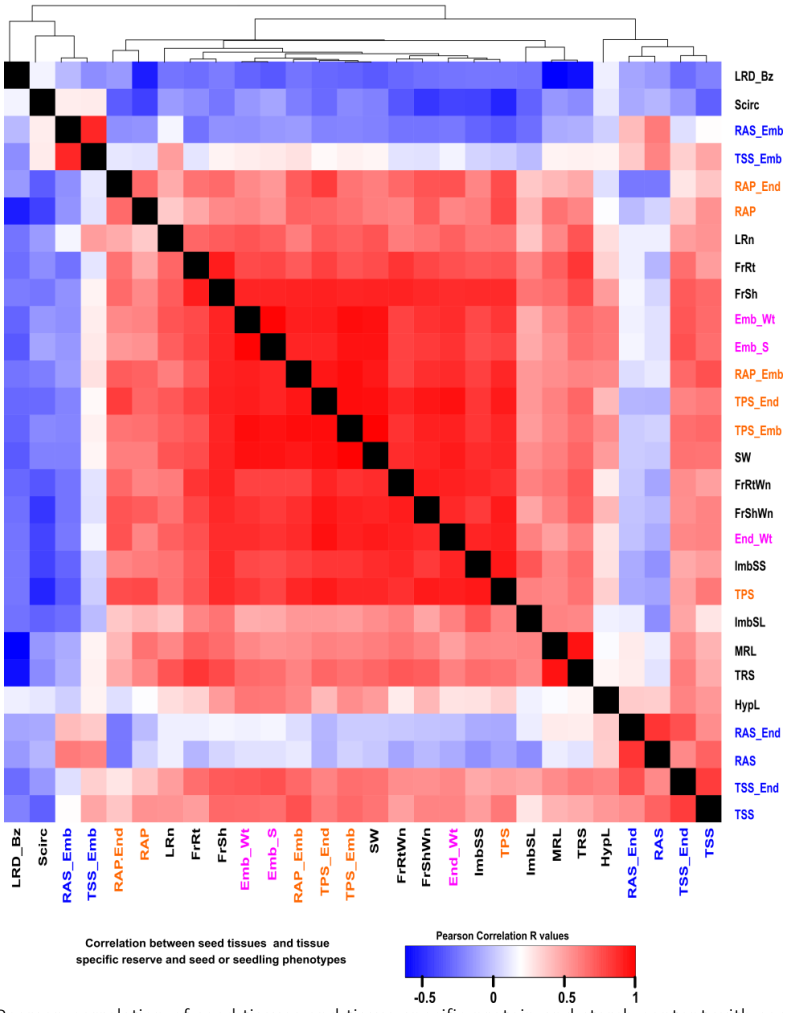


Figure 4.5. Pearson correlation of seed tissues and tissue specific protein and starch content with seed, seedling and RSA traits. Emb_wt= Embryo weight (mg/seed), End_wt= Endosperm weight (mg/seed), Emb_S= average Embryo Size (mm), Emb_L= average Embryo Length (mm), RAP_Emb= average Relative Amount Protein in Embryo (mg/gram of embryo), RAP_End= average Relative Amount of Protein in Endosperm (mg/gram of endosperm), TPS_Emb= average Total amount of Protein per embryo (µg/embryo), TPS_End= average Total amount of Protein per endosperm (µg/endosperm), RAS_Emb= average Relative Amount of Starch in Embryo (mg/gram of embryo), RAS_End= average Relative Amount of Starch in Endosperm (mg/gram endosperm), TSS_Emb = average Total amount of Starch per Embryo (µg/embryo), TSS_End = average Total amount of Starch per Endosperm (µg/endosperm), RAP= relative amount of protein (mg/gram seed), TPS= total amount of protein (µg/seed), RAS= relative amount of starch (mg/gram seed) and TSS= total amount of starch (µg/seed), SW= Seed Weight, SL= Seed Length, FrRt= Fresh Root weight, FrSh= Fresh Shoot weight, FrShWn= Fresh Shoot weight without nutrients, and FrRtWn= Fresh Root weight without nutrients.

The power of the S. lycopersicum x *S. pimpinellifolium RIL population for detection of QTLs for seed protein and starch content*

The power of QTL detection in a RIL population depends on several factors, including the number of RILs in the population, the magnitude of effects relative to the environmental and error variation, the heritability (h^2_b) of the trait, gene action, marker coverage, the distance between marker loci, linkage of QTLs and also the statistical techniques applied (Alonso-Blanco et al., 1998; Mackay, 2001; Foolad et al., 2003; Mackay et al., 2009). The variation and number of the RILs and the overall heritability of the trait (i.e. heritability in the broad sense) are the major players that strongly affect the quality of QTL analysis, including the LOD score values, explained variance, the number of QTLs detected and the accuracy of their map positions (Alonso-Blanco et al., 1998; Alonso-Blanco and Koornneef, 2000). However, the heritability in the broad sense can be increased by accurate determination of the phenotypes and eliminating the environmental factors affecting the traits (Ukai, 2000). We have utilized homogenous and strictly controlled plant growth conditions and reserve food phenotype testing and we employed a higher number of RILs in this study than in our previous ones, 83 vs. 100 (Kazmi et al., 2012; Khan et al., 2012). These contributed to a high broad sense heritability of the reserve food phenotypes (h^2_b 0.78-0.98, Table 4.2). The phenotypic variation of the parents, as well as in the RIL population and the resolution and size of this population were sufficient to find 24 QTLs for seed protein and starch content, demonstrating that this RIL population is a powerful tool for unravelling the quantitative traits under study. We observed a significant increase in the LOD scores, explained variance of individual and total explained variance of all the QTLs for a particular trait whereas the number of QTLs for the 18 previously described seed and seedling traits increased from 62 (Khan et al., 2012) to 108 QTLs (Figure 4.3 and Supplemental Table S4.2). Most of the QTLs that we had identified in the 83 lines in our previous study were recovered in the present analysis and, in addition, we found one or more extra QTLs per trait. This is in agreement with Alonso-Blanco et al. (1998) who reported that the power of detection of QTLs in a RIL population mainly depends on the number of the RILs analysed.

Physiological variation in whole seed protein and starch content and their correlation with other traits

The storage compounds that mostly accumulate during the seed filling phase are storage proteins (e.g., albumins, globulins and prolamins), oil (often triacylglycerols) and carbohydrates (often starch) (Baud et al., 2002). These reserves are of major importance as they support early seedling growth when degraded upon germination and, therefore, participate in crop establishment. The balance between protein and starch is genetically

regulated in different species and depends on the availability of nitrogen and carbon to the mother plant. For example, in most dicots there is higher accumulation of protein, mostly in the embryo, as compared to starch whereas in many monocots, e.g., cereals, there is higher accumulation of starch, mainly in the endosperm, as compared to protein (Bewley et al., 2012). In our present study we observed about 12 times more relative protein content (RIL mean= 86.51±10.15 mg/g seed as compared to starch (RIL mean= 6.85± 3.99 mg/g seed). The same was true for the total amount of protein per seed (RIL mean= 148.66±41.2 µg/seed) as compared to starch (RIL mean= 11.75±7.67 µg/seed). Of the two parents, *S. lycopersicum* had a 19 times higher relative protein content (mean= 99.23±2.79 mg/g) as compared to starch (mean= 5.16±0.76 mg/g) and *S. pimpinellifolium* a 12 times higher level of relative protein content (mean=71.76±2.38 mg/g) as compared to starch (mean= 5.69±0.76 mg/g).

The total amount of protein and starch per seed of the parents followed the same pattern as the relative amounts (Figure 4.1 and Table 4.1). In other studies in tomato (Schauer et al., 2005) also significantly higher levels of protein were detected as compared to starch in five wild tomato species as well as *S. lycopersicum*. In addition, significantly higher protein and protein-related compounds were found in leaves and fruits of *S. lycopersicum* than in *S. pimpinellifolium.* Contrarily, starch and starch-related compounds in the leaves and fruits of *S. pimpinellifolium* were higher than in *S. lycopersicum*. Similar results have been documented by (Toubiana et al., 2012) who observed significantly higher levels of protein-related metabolites in fruits and dry seeds of the *S. lycopersicum* M82 cultivar, as compared to most of the introgression lines (ILs) carrying a small chromosomal portion of the wild tomato species *S. pennellii* within the chromosomal background of M82. However, metabolites related to starch content were significantly higher in the ILs. Higher level of protein (25-39.9%) and minimal amounts of starch (0.5-0.8%) have been documented in different genotypes of Arabidopsis (Siloto et al., 2006), as well as in other dicots such as pea, common bean, chickpea and lentil (de Almeida Costa et al., 2006). Contrarily, in monocots, such as maize (Singletary and Below, 1989), barley (Ciulca et al., 2009) and rice (Mahesh et al., 2012) significantly greater amounts of starch were observed as compared to protein.

Good and uniform seedling emergence and vigour and the subsequent establishment of a vigorous seedling are crucial for sustainable and profitable crop production and therefore the seedling stage is considered the most critical stage of a developing crop. Seed vigour is one of the most important factors determining both seed germination and subsequent post-germination reserve-dependent seedling growth leading to the establishment of the seedling (Whittington, 1973; Hodgkin and Hegarty, 1978; Perry, 1984; Finch-Savage, 1995). Seed weight is an indication of the amounts of reserves that seeds may contain and large seeds establish vigorous seedlings supported by the larger amounts of reserve food that heavy seeds contain (Wright and Westoby, 1999). In the

present study we observed significantly higher levels of both RAP (99.23 mg/g) and TPS (291.04 µg/seed) in the seeds of the *S. lycopersicum* parent which is nearly three folds larger as compared to the RAP (71.76 mg/g) and TPS (77.91) obtained from the *S. pimpinellifolium* parent. Similar trends were revealed in the RIL population where large seeds had higher and small seeds had lower levels of both RAP and TPS (Figure 4.1, Table 4.1).

These results corroborate a common genetic basis underlying seed mass and seed protein content. The RAS of the two parents was statistically not different although there was greater variation in the RAS of the RIL population, indicating the control of seed starch content by other genetic factors than the ones that control seed mass. Nevertheless, the TSS of the *S. lycopersicum* parent was significantly higher than that of *S. pimpinellifolium,* revealing a relationship between total amount of reserve and seed mass. The RAP of the whole seed of the RIL population was significantly correlated with seed weight (FDR p value = 0.01) and strongly correlated with seedling traits under nutrientless conditions (Khan et al., 2012) with the FDR p-value ranging from 0.03 to 0.001). The RAP of the RIL population was also positively and strongly correlated (FDR p-value < 0.02) with seedling root system architectural (RSA) traits grown on half MS agar medium (Figure 4.2 and Supplemental Table S4.1). These particularly included main root path length (MRL) and total root size (TRS). However, no correlation was observed between the RAP and seedling fresh and dry shoot and root weight grown under normal nutritional conditions. Nonetheless, the TPS correlated strongly (FDR p-value < 0.00001) with all seed and seedling traits under both normal and nutrientless conditions, except hypocotyl length (HypL) and lateral root density per basal zone (LRD_Bz). Furthermore, it was significantly negatively correlated with seed circularity (SC). These results suggest that seed protein content strongly improves seedling vigour under stress conditions.

These results are in agreement with our previous finding (Khan et al., 2012) in which we found stronger correlation between seed weight and seedling biomass grown under nutrientless conditions compared to the correlation observed with seedling biomass grown with normal nutrition. In our previous study of seedling growth under nutrientless conditions, we also observed a higher decrease in the shoot and root weights of the small seed related RILs, including the *S. pimpinellifolium* parent, compared to the RILs with large seeds, including *S. lycopersicum.* These results are in agreement with those reported by Nieuwhof et al. (1989) who found significant correlation between tomato seed weight and seedling biomass grown under nutrient-deprived conditions. There were significant differences in the RAP, but the correlation analysis revealed that these differences in protein content were not sufficient to affect seedling vigour under normal nutrition. On the other hand, the variation in TPS was big and that might affect seedling vigour even under normal nutrient conditions and therefore, we observed a strong correlation between TPS and seedling biomass grown under normal nutrition. This is also obvious from the RAP and

TPS of the two parents with *S. lycopersicum* having only 29% higher RAP than *S. pimpinellifolium*. However, this difference jumps to nearly four-fold in case of TPS of the two parents due to differences in seed size (Figure 4.1, Table 4.1). This also supports our previous suggestion that positive effects of heavy seeds on seedling growth could be due to the high quantity of reserve food in larger seeds as compared to small seeds. The negative correlation between TPS and seed circularity (SC) is logical as higher values of TPS means higher embryo size and higher embryo length as TPS is strongly correlated with embryo traits. Thus theoretically the higher the TPS, the higher will be the embryo length and subsequently SL and the lower will be SC. Thus there is inverse relationship between TPS and SC. In addition, as in our RIL population higher SC values are related to small seed size and small and circular seeds might have less potential to store higher level of protein.

Seed size in different species is positively correlated with the protein content of the seed and, in turn, seed protein content is positively correlated with seedling vigour (Lowe and Ries, 1973; Ries and Everson, 1973; Evans and Bhatt, 1977; Saxena et al., 1987; Panthee et al., 2005). It implies that large-sized seeds will have a higher relative and total amount of protein and will produce more vigorous seedlings, as well as higher yield. In addition to genotype, the production and quality of protein is also affected by the environment. Ries and Everson (1973) reported that protein content of wheat seeds collected from five cultivar trials grown at several different locations were affected by both genotype and environment. The larger seed of 21 of the 25 cultivars contained significantly higher relative amounts of protein (mg protein/g of seed) than the smaller seeds and the protein content was significantly correlated with seed size and this correlation was very strong between total amount of protein per seed and seed size. Irrespective of genotype or environment, seedling vigour (expressed as dry weight of the shoot) was consistently related to seed protein content. Seed size also displayed a strong correlation with seedling vigour related traits, but when seed size was eliminated by comparing uniformly sized seeds, the relationship between seedling vigour and relative amount of protein was still significant. The best relationship was between milligrams of protein per seed and seedling weight, which indicates a consistent link between seed size, protein content and seedling vigour.

Unlike RAP the RAS of seed was not correlated with any seed or seedling trait. This could be due to several reasons. One of the reasons could be the non-consistent link between seed size and RAS level in the RIL population, as compared to RAP where there is linear increase in its level with increase in seed size. For example, the concentration of starch may be higher, but the total amount of starch as well as protein may be low if seed size is small and vice versa. In addition, no inverse relationship was observed between starch and protein content of the seed. Most studies carried out on the genetic control of fruit size and its composition show that most of the wild type tomato species have smaller fruit sizes and higher levels of soluble solid content, including starch. Most of these studies

have revealed negative correlations between fruit weight/size and fruit size (starch, brix and different sugars). Importantly, seed size/weight has been reported to be positively correlated with fruit size/weight and negatively with fruit soluble solid contents (Goldman et al., 1995; Fulton et al., 1997; Fulton et al., 2000; Monforte and Tanksley, 2000; Frary et al., 2004; Schauer et al., 2005; Prudent et al., 2009). Nevertheless the TSS showed significantly higher correlations with most of the seed and seedling biomass traits (FDR p-values ranging from 0.007 to 0.00001) but did not correlate with all the RSA traits and HypL (Figure 4.2, Supplemental Table S4.1). This suggests that, although the concentration of starch is unrelated with seed size, the total amount of starch per seed is still higher in large seeds and, might be sufficient to affect seedling growth. Positive correlations between whole rosette leaves and root starch content and seedling traits have been reported in Arabidopsis (El-Lithy et al., 2010), seed starch content and seed weight in maize (Singletary and Below, 1989) and seed starch content with seed and seedling traits in rice (Cui et al., 2002).

Genetic regulation of seed protein and starch content and their interaction with seed and seedling traits

The genetic variation existing in the Solanum species for seed quality has hardly been employed in breeding programs. Although extensive information about the genetic control of seed quality in tomato and its relation with seedling vigour has been provided in previous studies and QTLs for seed weight or size have been identified, none of these discuss in detail the genetic control of seed food reserves and their relationship with seed quality and seedling establishment (Goldman et al., 1995; Grandillo and Tanksley, 1996; Doganlar et al., 2000; Tanksley, 2004; Khan et al., 2012). In our current study we identified 24 putative genomic regions regulating the four studied reserve food related traits (RAP, TPS, RAS and TSS). One major and one minor QTL was detected for RAP with total explained variance of 29.2%. The additive effect for the major QTL was from *S. lycopersicum* parent while for the minor QTL was from *S. pimpinellifolium* parent which is in agreement to the relative amount of protein of the two parents. The major RAP QTL is co-locating with a cluster of 14 seed and seedling weight and RSA traits QTLs identified in our previous study (Fig 4.3 A and B, Table 4.2 and Supporting Information Table 4.2). This is also in agreement with the moderate correlation of RAP with most seed and seedling traits. However, a higher number of QTLs (6 QTLs) was detected for TPS which highly explained the trait (53.1%). For all the 6 QTLs for TPS the alleles of *S. lycopersicum* parent increased the trait. This is in agreement with our previous finding (Khan et al 2012) where for all the QTLs related to seed weight the favourable alleles were derived from the *S. lycopersicum* parent (Supporting Information Table 4.2) and reinforce the concept that large seeds have higher amount of total reserve food compared to small seeds. These findings are also in

agreement with Doganlar et al. (2002) and Ashrafi et al. (2012), who detected QTLs for seed and fruit size in a RIL population obtained from a cross between *S. lycopersicum* X *S. pimpinellifolium* and observed that all the seed and fruit size QTLs were contributed by the positive alleles of the cultivated tomato (*S. lycopersicum*). Importantly, all the 6 TPS QTLs are co-locating with SW QTLs as well as several seedling vigour and RSA traits related QTLs (Figure 4.3 A,B, Table 4.2 and Supplemental Table S4.2). 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 genetic mechanistic basis or a dependency of traits (Clerkx et al., 2004).

Thus, the co-location of QTLs for RAP and TPS with seed weight and seedling vigour QTLs with same direction of additive effect and the strong correlation in the phenotypic values suggest pleiotropic interactions among these traits. Co-location of QTLs for tomato seed weight with several QTLs for soluble solids and fruit weight on chromosome 9 have been reported in previous studies (Goldman et al., 1995; Doganlar et al., 2000). Co-location of QTLs between seed-weight and other life traits have been documented by Alonso-Blanco et al. (1999) in *Arabidopsis*. Association of QTLs for protein content (%), 1000 grains weight and endosperm hardness have been reported in barley (Walker et al., 2011), for protein content, 1000 grain weight and yield in wheat (Groos et al., 2003), for seed protein, seed and plant traits in pea (Burstin et al., 2007) and for grain protein and plant morphological traits in sorghum (Rami et al., 1998).

Contrary to RAP, we identified a higher number of QTLs accompanied by larger total explained variance for TPS as well as RAS (7 QTLs) and TSS (9 QTLs) having total explained variance of 53.9 and 66.5% respectively. Although, the phenotypic values of the two parents for RAS were statistically undistinguishable, we detected 7 significant QTLs on 6 different chromosomes with total explained variance of 53.8%. This was the results of a strong transgression in the RIL population for this trait. Out of the 7 QTLs identified for RAS, four QTLs were enhanced by a positive allelic effect of the *S. lycopersicum* parent and three QTLs were enhanced by the *S. pimpinellifolium* parent. Four of the RAS QTLs are also co-locating with either both seed and seedling traits ($RAS_{1\text{-}2}$ and RAS_7) or only a single seedling trait QTL ($RAS_{2\text{-}1}$ and RAS_4), whereas 3 of the RAS QTLs ($RAS_{2\text{-}2}$, RAS_{11} and RAS_{12}) are specific QTLs that are not co-locating with any other trait. This might indicate that the concentration of starch in the seed is partially related with seed weight, but is mostly controlled by genetic processes different than the one that control seed size. This was also expected from the non-significant correlation between phenotypic values for RAS and seed and seedling traits. Although *S. lycopersicum* has large seed size and has significantly higher amount of TSS compared to *S. pimpinellifolium*, 4 TSS QTLs were increased by the positive alleles of the *S. lycopersicum* parent, whereas 5 TSS QTLs were enhanced by the *S. pimpinellifolium* parent. Both the LOD score and explained variance of those TSS QTLs co-locating with seed weight ($TSS_{1\text{-}2}$ and TSS_7) became doubled while those not co-locating seed weight decreased significantly (Figure 4.3 and Table 4.2). Seven of the 9 TSS QTLs are

also co-locating with either both seed and seedling traits or only seedling traits (Figure 4.3 and Table 4.2). This increase in number of co-locating QTLs as well as in power of TSS QTLs co-locating with seed weight resulted in moderate but highly significant positive correlation of TSS with most seed and several seedling vigour related traits (Figure 4.2, Supplemental Table S4.1). The sharing of QTLs for RAP, and TPS with the QTLs for seed and seedling biomass and RSA vigour related traits, with the positive additive effects from the same parent and the positive correlation among the phenotypic values for these traits, suggest a strong link between seed protein and seed and seedling biomass (Sun et al., 2008). On the other hand sharing of small number of RAS and TSS QTLs with seed and seedling trait reveal partial dependency of starch concentration on seed size. Further, the absence of both co-location and correlation between RAP and RAS indicates that the seed protein and starch content are completely independent of each other and their biosynthesis and accumulation in the seed is controlled by different regulatory genes.

The interaction between embryo, endosperm and seed quality and seedling vigour

Seed size or weight can putatively be affected by the genotype of three different seed tissues of tomato: the maternal testa, the triploid endosperm and the diploid embryo (Doganlar et al., 2000; Orsi and Tanksley, 2009). We observed that 70-80% of the seed size is explained by embryo size whereas the endosperm and testa are relatively thin and contribute less to seed size. However, testa and endosperm together contribute 65-70% to the seed weight and the embryo 30-40%. This ratio of embryo weight to seed weight is in the same range as reported by Sheoran et al. (2005), who reported that in *S. lycopersicum* seeds the embryo accounts for about 35% of the total seed weight. Irrespective of the ratio, both the embryo and endosperm weight of *S. lycopersicum* and the heavier 10% RILs were significantly higher than *S. pimpinellifolium* as well as the lighter 10% RILs of the population (Figure 4.4, Table 4.3 and Supplemental Table S4.3). In addition, embryo and endosperm weight and size were strongly correlated with seed dimension traits (FDR p value < 0.00001) and these seed tissues were also strongly correlated with each other (FDR p value <0.00001). This suggests that in tomato, seed size depends on both the embryo and endosperm tissues, which might be genetically interlinked. This observation is in agreement with Zhang and Maun (1993), who observed similar correlations within embryo and endosperm masses and between seed mass and dimensions in *Calamovilfa longifolia*. In a number of cultivars of field-grown barley, it was found that endosperm mass was positively correlated with 1000 grain weight and dry grain size (Cochrane and Duffus, 1983). In three strains of subterranean clover, embryo weight accounted for about 66% of the seed weight and there was strong correlation (r^2=0.99) between seed and embryo weight (Black, 1957). Correlations between embryo cell number and seed dry weight have been reported for pea (Davies, 1975), soybean (Egli et al., 1981), maize (Reddy and Daynard, 1983) and wheat (Jenner et al., 1991).

Embryo and endosperm are genetically different tissues. In many cases different genes or differential expression of genes have been reported (Penfield et al., 2006), but there is evidence in support to our finding that tightly links embryo, endosperm and seed size with each other. For example Orsi and Tanksley (2009), found that an ABC transporter controls seed size in tomato through gene expression in the developing embryo. They further revealed that there was a concomitant increase in seed, embryo and endosperm size during different seed developmental stages but there was no difference in the ratio between embryo and endosperm of small and large seeded tomato lines. The relationship between embryo, endosperm and seed weight or size have been reported in other crops. In Arabidopsis Garcia et al. (2003), reported that *HAIKU1* (*IKU1*) and *IKU2* control seed size through their effect on endosperm and integument development. The *iku* mutants showed an early arrest of increase in seed size through premature arrest of endosperm proliferation; inhibited cell division in the embryo and also restricted cell elongation in the integument. Expression of *AP2* during Arabidopsis seed development decreased seed mass (Jofuku et al., 2005; Ohto et al., 2005). In the loss-of-function *ap2* mutants, seed mass was increased through an increase in cell division and enhancement of seed filling during the maturation period of the embryo. The endosperm showed a prolonged period of rapid growth accompanied by delayed cellularization in *ap2* mutants (Ohto et al., 2009). Thus, *AP2* regulated seed mass through its effect on both the embryo and the covering tissues.

In addition to seed dimension, we observed strong correlation (FDR p value < 0.001) between embryo weight and size, endosperm and seedling biomass under normal and nutritional stress conditions (FrRt, FrSh, FrRtWn, FrShWn) whereas the correlation of these seed tissues was lower but statistically significant with the RSA traits (MRL, TRS, LRn). This is logical as all these three tissues have a strong correlation with seed traits and seed traits have weak to medium correlation with RSA traits. Further, the correlation between embryo and seedling traits and endosperm and seedling traits was nearly the same. The only difference here was that both embryo weight and size correlated strongly with hypocotyl length (FDR p value 0.003) whereas the endosperm did not correlate with hypocotyl length. Obviously, the hypocotyl is part of the embryo and therefore may depend more on embryo size than the endosperm reserves. These results are in agreement with studies on *Calamovilfa longifolia* (Zhang and Maun, 1993) and on three strains of subterranean clover (Black, 1957). In wheat seedling vigour and seedling characteristics were highly correlated with embryo size (Richards and Lukacs, 2002). Embryo size seems to be a major determinant of early seedling vigour in maize and most temperate cereals and is also closely related with seed size (López-Castañeda et al., 1996). However, seedling mass was mainly correlated with endosperm size rather than the embryo in *Panicum virgatum, Agropyron psammophilus and* seven dune species (Zhang and Maun, 1989; Zhang and Maun, 1991). Again, this suggests that embryo and endosperm are closely interlinked seed traits.

Protein and starch content of seed tissues and their interaction with seed, seedling and RSA traits

With respect to seed quality the two seed tissues can be differentiated in terms of the kind and amount of the reserves that they accumulate. For instance, we observed significantly higher amounts of protein in embryo (RAP_Emb and TPS_Emb) as compared to the RAP and TPS of endosperm in both groups of the extreme RILs. In addition, the RAP of embryo and endosperm of the heavier 10% RILs were significantly lower (p value < 0.001) as compared to the smaller 10% RILs. However, there was no significant difference in the ratio of embryo and endosperm protein between the two groups. This means that the increase in protein content is linear with the increase in mass of seed tissues. In correspondence to the whole seed, we also quantified significantly lower levels of starch in both embryo and endosperm of both the parents as well as in the RIL groups. Dicots store higher amounts of protein mainly in the embryo rather than starch (Bewley et al., 2012). Our results suggest a stronger role for the embryo protein as well as embryo in seed mass accumulation and seedling establishment than the endosperm.

Other studies have also reported higher levels of both protein and oil to be related with the embryo and starch with the endosperm. In alfalfa, embryo size and vigour were positively related to the levels of storage protein, nitrogen and free amino acids (Lai and McKersie, 1994) whereas starch accumulation was inversely related to embryo weight and quality. In contrast, in maize the starch content was significantly higher as compared to protein and endosperm dry weight was highly correlated with the starch content (Singletary and Below, 1989).

Although the extreme RILs had significantly different (p value < 0.001) RAS and TSS of both embryo and endosperm, the group analysis revealed that there was no significant difference between the RAS of endosperm of the two groups due to the large variation within the RIL groups. However, when the RAS of the endosperm was calculated on a seed weight basis as TSS_End, then *S. lycopersicum* as well as the heavier 10% RILs became significantly higher (p value <0.001) in TSS_End compared to *S. pimpinellifolium,* as well as the smaller 10% of the extreme RILs. This indicates that although the relative amount is variable within the large and small seeded groups the total amount of starch per seed is greater in the larger seeds. This is also in agreement with the data for whole seed starch content where we observed greater variation in the RAS of the RIL population, irrespective of seed size. The RAS of the two parents was statistically indistinguishable with the result that the TSS of *S. lycopersicum* was significantly higher than that of *S. pimpinellifolium.* In contrast, the RAS of the embryo of *S. pimpinellifolium* was significantly higher (p value< 0.0001) than the RAS of the embryo of *S. lycopersicum* as well as the RAS_Emb of the smaller 10% RILs were higher (p value 0.03) than that of the heavier 10% RILs. This suggests that the small seeds accumulate significantly higher relative amounts of starch in the embryo as compared to endosperm. Further, as in the case of whole seed RAS, the RAS of

both the embryo and endosperm did not correlate with any single seed or seedling trait. Compared to the correlation with the RAS of the whole seed, the RAS of the embryo was not correlated with the TSS of the whole seed whereas the RAS of the endosperm was strongly correlated with the TSS of the whole seed. This indicates that the endosperm accounts for the maximal portion of whole seed starch content. Further the TSS of the embryo was not correlated with any seed or seedling trait while the TSS of endosperm was strongly correlated with most of the seed and seedling traits. This correlation between TSS and seed and seedling traits appears to be contributed by the seed weight variation in the RILs lines and is also consistent with the correlation between TSS of whole seed and seed and seedling biomass.

Conclusion

In conclusion, we have accepted the hypothesis that large seeds have higher amount of reserve food in respect to protein and the higher amount of protein support the establishment and early vigour of seedling. Since tomato seeds also contain high level of lipid (20-29%) (Sheoran et al., 2005) and we didn't study lipid content, we don't know how this lipid content relates with seed size and what is its correlation with seed starch and protein content is unknown. The current study reveals that RAP and TPS have strong correlation with seed and seedling vigour related traits and a pleiotropic association with SW QTLs, hence occupy an important position in seed weight and seedling establishment. In case of starch there is no clear link between RAS and seed size, although there is a highly significant correlation between TSS and seed and most seedling traits and several RAS and TSS QTLs are co-locating with seed and seedling biomass QTLs. This indicates that irrespective of the level of starch (RAS), the total amount of starch (TSS) of large seeds is higher compared to small seeds. Furthermore, a strong correlation of both embryo and endosperm weight was observed with seed and seedling biomass. This implies that seed size and embryo and endosperm size are under the control of common genetic mechanisms. However, the embryo traits (Emb_Wt and Emb_S) were relatively stronger associated with seed and most seedling traits compared to endosperm. In addition the protein content of embryo was significantly higher than the endosperm and was strongly correlated with all seed, seedling and RSA traits including hypocotyl length, where the endosperm protein content was not correlated with RSA and hypocotyl length (Supplemental Table S4.3). These results indicate that the embryo is more associated with reserve accumulation and in supporting early seedling growth in tomato.

Both the strong correlation and co-location of QTLs for seed reserve with the QTLs for seed and seedling biomass with the same direction of the genetic effect indicate a strong association between seed reserve food and seed quality and seedling biomass. Our

results can be used in marker-assisted selection for improved seed reserve food and, hence, seed quality and seedling or crop establishment. Furthermore, our study contributes to the generation of new testable hypotheses and may expand our fundamental understanding of reserve food behavior affected by genetic background. The genomic regions identified will help in cloning the causal genes especially those with moderate to high broad sense heritabilities (Dudley, 1993; Tanksley, 1993). Molecular markers linked to the QTLs may providing a rapid method to select for specific genotypes without the need to extensively assess phenotypes at all stages in the breeding program.

Supporting information

Supporting information can be downloaded from
http://www.wageningenseedlab.nl/thesis/nkhan/SI/chapter4

Supplemental Table S4.1. Correlation between whole seed reserve food and seed and seedling phenotypes. RAP= Relative Amount of Protein (mg/g of seed), TPS= Total amount of Protein (µg/seed) , RAS= Relative Amount of Starch (mg/g of seed), TSS= Total amount of Starch (µg/seed), FrRt= Fresh Root weight, FrSh= Fresh Shoot weight, HypL= Hypocotyl Length, MRL= Main Root Length, TRS= Total Root Size, LRn= Lateral Root number, LRD_Bz= Lateral Root Density per Branch zone, SW= Seed Weight, SS= Seed Size, SL= Seed Length, SC= Seed Circularity, ImbSS= Imbibed Seed Size, ImbSL= Imbibed Seed Length, FrRtWn= Fresh Root weight under nutrientless conditions, DrRtWn= Dry Root weight under nutrientless conditions, FrShWn= Fresh Shoot weight under nutrientless conditions, DrShWn= Dry Shoot weight under nutrientless conditions.

Supplemental Table S4.2. Overview of significant QTLs associated with seed and seedling traits of the *S. lycopersicum* x *S. pimpinellifolium* RIL population (100 Lines).

1. Traits: 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 conditions, FrRtWn= Fresh Root weight under nutrientless conditions, , DrShWn= Dry Shoot weight under nutrientless conditions, DrRtWn= Dry Root weight under nutrientless conditions, MRL= Main Root Length, TRS=Total Root Size, LRn= Lateral Root number, LRD_Bz= Lateral Roots Density per Branched zone.
2. Chromosomes on which the QTLs were detected.
3. 1-LOD support interval in centi-Morgan.
4. Nearest marker to the position of ìdentified QTL.
5. LOD score (LOD score of 2 or above was calculated to be significant for this population)

6. Additive effect; 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.
7. Percentage of variation explained by each QTL.
8. Percentage of total variation explained by genetic factors for a single trait as estimated by MapQTL.
9. Broad-sense heritability estimate for each trait.

Supplemental Table S4.3. Correlation between seed tissue specific reserve food and seed and seedling phenotypes. Endo_Wt= Endosperm Weight per seed, Emb_Wt= Embryo Weight per seed, Embr_S= Embryo Size, RAP_End= Relative Amount of Endosperm Protein (mg/g of seed), RAP_Emb= Relative Amount of Embryo Protein (mg/g of seed), TPS_End= Total amount of endosperm protein (µg/seed), TPS_Emb= Total amount of Embryo Protein (µg/seed), RAS_End= Relative Amount of Endosperm Starch (mg/g of seed), RAS_Emb= Relative Amount of Embryo Starch (mg/g of seed), TSS_End= Total amount of Endosperm Starch (µg/seed), TSS_Emb= Total amount of Embryo Starch (µg/seed), FrRt= Fresh Root weight, FrSh= Fresh Shoot weight, HypL= Hypocotyl Length, MRL= Main Root Length, TRS= Total Root Size, LRn= Lateral Root number, LRD_Bz= Lateral Root density per Basal zone, SW= Seed Weight, SC= Seed circularity, ImbSS= Imbibed seed size, ImbSL= Imbibed Seed Length, FrRtWn= Fresh Root weight under nutrientless conditions, FrShWn= Fresh Shoot weight under nutrientless conditions, WRAP= Relative Amount of Whole seed Protein (mg/g of seed), WTPS= Total amount of Whole seed Protein (µg/seed), WRAS= Relative Amount of Whole seed Starch (mg/g seed), WTSS= Total amount of Whole seed Starch (µg/seed).

References

Alonso-Blanco C, Blankestijn-de Vries H, Hanhart C, Koornneef M (1999) Natural allelic variation at seed size loci in relation to other life history traits of Arabidopsis thaliana. Proceedings of the National Academy of Sciences of the United States of America **96:** 4710

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

Alonso-Blanco C, Koornneef M, Stam P (1998) The use of recombinant inbred lines (RILs) for genetic mapping. Methods in molecular biology **82:** 137-146

Ashrafi H, Kinkade MP, Merk HL, Foolad MR (2012) Identification of novel quantitative trait loci for increased lycopene content and other fruit quality traits in a tomato recombinant inbred line population. Molecular Breeding **30:** 549-567

Barlow E, Donovan G, Lee J (1983) Water relations and composition of wheat ears grown in liquid culture: effect of carbon and nitrogen. Functional Plant Biology **10:** 99-108

Baud S, Boutin JP, Miquel M, Lepiniec L, Rochat C (2002) An integrated overview of seed development in Arabidopsis thaliana ecotype WS. Plant Physiology and Biochemistry **40:** 151-160

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. Annals of statistics **29:** 1165-1188

Bentsink L, Alonso-Blanco C, Vreugdenhil D, Tesnier K, Groot SP, Koornneef M (2000) Genetic analysis of seed-soluble oligosaccharides in relation to seed storability of Arabidopsis. Plant Physiology **124:** 1595-1604

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

Bettey M, Finch-Savage W, King G, Lynn J (2008) Quantitative genetic analysis of seed vigour and pre-emergence seedling growth traits in Brassica oleracea. New Phytologist **148:** 277-286

Bewley JD, Bradford KJ, Hilhorst HWM, Nonogaki H (2012) Seeds: Physiology of Development, Germination and Dormancy. Springer Verlag

Black J (1957) The early vegetative growth of three strains of subterranean clover (Trifolium subterraneum L.) in relation to size of seed. Crop and Pasture Science **8:** 1-14

Bonfil C (1998) The effects of seed size, cotyledon reserves, and herbivory on seedling survival and growth in Quercus rugosa and Q. laurina (Fagaceae). American Journal of Botany **85:** 79-79

Burstin J, Marget P, Huart M, Moessner A, Mangin B, Duchene C, Desprez B, Munier-Jolain N, Duc G (2007) Developmental genes have pleiotropic effects on plant morphology and source capacity, eventually impacting on seed protein content and productivity in pea. Plant Physiology **144:** 768-781

Ciulca A, Madosa E, Ciulca S, Friskan I (2009) Evaluation of breeding potential for some quality traits in winter barley using multiple selection indices and multivariate analysis. Journal of Horticulture, Forestry and Biotechnology **13:** 287-291

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

Cui K, Peng S, Xing Y, Xu C, Yu S, Zhang Q (2002) Molecular dissection of seedling-vigor and associated physiological traits in rice. Theoretical and Applied Genetics **105:** 745-753

de Almeida Costa GE, da Silva Queiroz-Monici K, Pissini Machado Reis SM, de Oliveira AC (2006) Chemical composition, dietary fibre and resistant starch contents of raw and cooked pea, common bean, chickpea and lentil legumes. Food Chemistry **94:** 327-330

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

Donovan GR, Lee J (1977) The growth of detached wheat heads in liquid culture. Plant Science Letters **9:** 107-113

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

El-Lithy ME, Reymond M, Stich B, Koornneef M, Vreugdenhil D (2010) Relation among plant growth, carbohydrates and flowering time in the Arabidopsis Landsberg erecta× Kondara recombinant inbred line population. Plant, Cell & Environment **33:** 1369-1382

Evans L, Bhatt G (1977) Influence of seed size, protein content and cultivar on early seedling viour in wheat.. Canadian Journal of Plant Science **57:** 929-935

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 **179**: 582–589

Foolad M, Zhang L, Subbiah P (2003) Genetics of drought tolerance during seed germination in tomato: inheritance and QTL mapping. Genome **46:** 536-545

Frary A, Fulton T, Zamir D, Tanksley S (2004) Advanced backcross QTL analysis of a *Lycopersicon esculentum* × *L. pennellii* cross and identification of possible orthologs in the Solanaceae. Theoretical and Applied Genetics **108:** 485-496

Fulton T, Beck-Bunn T, Emmatty D, Eshed Y, Lopez J, Petiard V, Uhlig J, Zamir D, Tanksley S (1997) QTL analysis of an advanced backcross of *Lycopersicon peruvianum* to the cultivated tomato and comparisons with QTLs found in other wild species. Theoretical and Applied Genetics **95:** 881-894

Fulton T, Grandillo S, Beck-Bunn T, Fridman E, Frampton A, Lopez J, Petiard V, Uhlig J, Zamir D, Tanksley S (2000) Advanced backcross QTL analysis of a *Lycopersicon esculentum* × *Lycopersicon parviflorum* cross. Theoretical and Applied Genetics **100:** 1025-1042

Garcia D, Saingery V, Chambrier P, Mayer U, Jürgens G, Berger F (2003) Arabidopsis haiku mutants reveal new controls of seed size by endosperm. Plant Physiology **131:** 1661-1670

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

Groos C, Robert N, Bervas E, Charmet G (2003) Genetic analysis of grain protein-content, grain yield and thousand-kernel weight in bread wheat. Theoretical and Applied Genetics **106:** 1032-1040

Hodgkin T, Hegarty T (1978) Genetically determined variation in seed germination and field emergence of *Brassica oleracea*. Annals of Applied Biology **88:** 407-413

Jansen R, 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

Jofuku KD, Omidyar PK, Gee Z, Okamuro JK (2005) Control of seed mass and seed yield by the floral homeotic gene APETALA2. Proceedings of the National Academy of Sciences of the United States of America **102:** 3117-3122

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. Plant Journal **62:** 148-159

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

Kehui C, Shaobing P, Yongzhong X, Sibin Y, Caiguo X (2002) Molecular dissection of relationship between seedling characteristics and seed size in rice. Acta Botanica Sinica **44:** 702-707

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. PlosOne **7:** e43991

Kidson R, Westoby M (2000) Seed mass and seedling dimensions in relation to seedling establishment. Oecologia **125:** 11-17

Lai FM, McKersie BD (1994) Regulation of starch and protein accumulation in alfalfa (*Medicago sativa* L.) somatic embryos. Plant Science **100:** 211-219

Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics **121:** 185-199

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

Lohaus G, Moellers C (2000) Phloem transport of amino acids in two *Brassica napus* L. genotypes and one B. carinata genotype in relation to their seed protein content. Planta **211:** 833-840

López-Castañeda C, Richards R, Farquhar G, Williamson R (1996) Seed and seedling characteristics contributing to variation in early vigor among temperate cereals. Crop Science **36:** 1257-1266

Lowe L, Ries SK (1973) Endosperm protein of wheat seed as a determinant of seedling growth. Plant Physiology **51:** 57-60

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

Maes T, Van de Steene N, Zethof J, Karimi M, D'Hauw M, Mares G, Van Montagu M, Gerats T (2001) Petunia Ap2-like genes and their role in flower and seed development. Plant Cell **13:** 229-244

Mahesh S, Bhattacharya A, Ghosh M, Chakarvarty A, Pal S (2012) Influence of flag leaf clipping, variety and sowing dates on nutritional quality in rice [*Oryza sativa* L.]. Journal of Crop and Weed **8:** 173-177

Monforte A, Tanksley S (2000) Fine mapping of a quantitative trait locus (QTL) from *Lycopersicon hirsutum* chromosome 1 affecting fruit characteristics and agronomic traits: breaking linkage among QTLs affecting different traits and dissection of heterosis for yield. Theoretical and Applied Genetics **100:** 471-479

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

Ohto M, Fischer RL, Goldberg RB, Nakamura K, Harada JJ (2005) Control of seed mass by APETALA2. Proceedings of the National Academy of Sciences of the United States of America **102:** 3123-3128

Ohto M, Floyd SK, Fischer RL, Goldberg RB, Harada JJ (2009) Effects of APETALA2 on embryo, endosperm, and seed coat development determine seed size in Arabidopsis. Sexual plant reproduction **22:** 277-289

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

Panthee D, Pantalone V, West D, Saxton A, Sams C (2005) Quantitative trait loci for seed protein and oil concentration, and seed size in soybean. Crop Science **45:** 2015-2022

Penfield S, Li Y, Gilday AD, Graham S, Graham IA (2006) Arabidopsis ABA INSENSITIVE4 regulates lipid mobilization in the embryo and reveals repression of seed germination by the endosperm. Plant Cell **18:** 1887-1899

Perry D (1984) Report of the vigour test committee, 1980-1983. Seed Science Technology **12:** 301-308

Poorter L, Rose SA (2005) Light-dependent changes in the relationship between seed mass and seedling traits: a meta-analysis for rain forest tree species. Oecologia **142:** 378-387

Prudent M, Causse M, Génard M, Tripodi P, Grandillo S, Bertin N (2009) Genetic and physiological analysis of tomato fruit weight and composition: influence of carbon availability on QTL detection. Journal of Experimental Botany **60:** 923-937

Rami JF, Dufour P, Trouche G, Fliedel G, Mestres C, Davrieux F, Blanchard P, Hamon P (1998) Quantitative trait loci for grain quality, productivity, morphological and agronomical traits in sorghum (*Sorghum bicolor* L. Moench). Theoretical and Applied Genetics **97:** 605-616

Richards R, Lukacs Z (2002) Seedling vigour in wheat-sources of variation for genetic and agronomic improvement. Crop and Pasture Science **53:** 41-50

Ries S, Everson E (1973) Protein content and seed size relationships with seedling vigor of wheat cultivars. Agronomy Journal **65:** 884-886

Roy S, Hamid A, Miah MG, Hashem A (1996) Seed size variation and its effects on germination and seedling vigour in rice. Journal of Agronomy and Crop Science **176:** 79-82

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

Saxena K, Faris D, Singh U, Kumar R (1987) Relationship between seed size and protein content in newly developed high protein lines of pigeonpea. Plant Foods for Human Nutrition (Formerly Qualitas Plantarum) **36:** 335-340

Schauer N, Zamir D, Fernie AR (2005) Metabolic profiling of leaves and fruit of wild species tomato: a survey of the *Solanum lycopersicum* complex. Journal of Experimental Botany **56:** 297-307

Sheoran IS, Olson DJH, Ross ARS, Sawhney VK (2005) Proteome analysis of embryo and endosperm from germinating tomato seeds. Proteomics **5:** 3752-3764

Siloto RMP, Findlay K, Lopez-Villalobos A, Yeung EC, Nykiforuk CL, Moloney MM (2006) The accumulation of oleosins determines the size of seed oilbodies in Arabidopsis. Plant Cell **18:** 1961-1974

Singletary GW, Below FE (1989) Growth and composition of maize kernels cultured in vitro with varying supplies of carbon and nitrogen. Plant Physiology **89:** 341-346

Sulpice R, Pyl ET, Ishihara H, Trenkamp S, Steinfath M, Witucka-Wall H, Gibon Y, Usadel B, Poree F, Piques MC (2009) Starch as a major integrator in the regulation of plant growth. Proceedings of the National Academy of Sciences of the United States of America **106:** 10348-10353

Sun H, Lü J, Fan Y, Zhao Y, Kong F, Li R, Wang H, Li S (2008) Quantitative trait loci (QTLs) for quality traits related to protein and starch in wheat. Progress in Natural Science **18:** 825-831

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

Tanksley SD (2004) The genetic, developmental, and molecular bases of fruit size and shape variation in tomato. Plant Cell **16:** S181-S189

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

Ukai Y (2000) Genetic analysis at the genomic level: Map and QTL. *In*. University of Tokyo Press, Tokyo, Japan

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

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

Walker CK, Panozzo J, Ford R, Eckermann P, Moody D, Lehmensiek A, Appels R (2011) Chromosomal loci associated with endosperm hardness in a malting barley cross. Theoretical and Applied Genetics **122:** 151-162

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

Wobus U, Weber H (1999) Seed maturation: genetic programmes and control signals. Current opinion in plant biology **2:** 33-38

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

Wurschum T, Groß-Hardt R, Laux T (2006) APETALA2 regulates the stem cell niche in the Arabidopsis shoot meristem. Science Signalling **18:** 295

Yin X, Struik PC, Van Eeuwijk FA, Stam P, Tang J (2005) QTL analysis and QTL-based prediction of flowering phenology in recombinant inbred lines of barley. Journal of Experimental Botany **56:** 967-976

Zhang J, Maun M (1993) Components of seed mass and their relationships to seedling size in *Calamovilfa longifolia*. Canadian Journal of Botany **71:** 551-557

Chapter 5

Canonical Association Reveals a Strong Link between Metabolic Signatures of Seed and Seedling Quality in a Recombinant Inbred Population of Tomato

Khan H, Willems LAJ, Ligterink W, Hilhorst HWM

Abstract

Plant growth and development are tightly linked to primary metabolism and are subject to natural variation, especially during early stages of development. In order to obtain knowledge of the genetic factors controlling seed and seedling biomass and primary metabolism, and to determine their relationship, a recombinant inbred lines (RIL) population of *Solanum lycopersicum* (cv. Money maker) x *Solanum pimpinellifolium* was analysed with respect to seed and seedling biomass and primary metabolite composition using a GC-TOF-MS metabolic profiling approach. This mQTL study was executed using a so-called generalized genetical genomics (GGG) approach in which for half of the lines of the population the dry seeds were profiled and for the other half 6h imbibed seeds. We could detect 160 metabolites and out of these 66 could be identified as known metabolites. A total of 112 genetic (G QTLs) and 34 loci controlling genetic-by-environment interactions (G x E QTLs) were identified for all the 160 metabolites. Many of these mQTLs coincided with seed weight and size, and seedling biomass QTLs, supporting the concept that metabolite profiles of seed and seedling biomass are genetically linked. A highly significant canonical correlation revealed that a specific combination of metabolites could explain the phenotype of seed and seedling. Especially the metabolites of 6 hour-imbibed seeds positively explained the phenotype of seed and seedling versus those of the dry seeds. This information can be used to build a comprehensive picture of associations between metabolites and seed and seedling phenotypes and provide a first step to the unravelling of the complex metabolic networks that influence seed and seedling quality.

Introduction

Plant growth and development are the result of interaction between genotype and the environment and are firmly associated with primary metabolism. Being an autotroph, the ability of a green plant to grow depends on its own photosynthetic and metabolic capacity. Therefore, the phenotype displayed by a plant and the biomass accumulated during the vegetative growth phase can be considered as the ultimate expression of its metabolic performance.

Plants function as integrated systems, in which metabolic and developmental pathways draw on common resource pools and respond to changes in environmental energy and resource supplies (Tonsor et al., 2004). Plants are rich in metabolites and ample variation exists in both the composition and content of these metabolites within plant species and most metabolites are poorly understood genetically (Keurentjes et al., 2006). This natural and genetic variation in metabolites and the subsequent expression of the complex phenotypic trait is due to the combined effects of multiple genes detectable as quantitative trait loci (QTLs) (Lisec et al., 2007; Keurentjes and Sulpice, 2009). Examples of multiple genes that govern the growth of multicellular organisms where each gene contributes a small share to the overall phenotype have been shown in mouse (Rocha et al., 2004), chicken (Jacobsson et al., 2005), Arabidopsis (El-Lithy et al., 2004) and rice (Li et al., 2006). While in plants several monogenic efforts targeting the production and/or distribution of primary metabolites within various part of the plant with the goal of modifying growth and/or biomass have been described, the success rate of these attempts has been rather limited (Lisec et al., 2007).

Quantitative trait analysis is the most common approach to the quantification of complex traits in a mapping population derived from a cross between distinct genotypes. In such analysis, genetic markers are associated with the phenotypic variation and are defined as quantitative trait loci (QTLs) (Liu, 1997; Lynch and Walsh, 1998). However, such traits are the end result of several intermediary biological steps (transcription, translation, post-translation, molecular response, etc.,) from genotype to phenotype (Keurentjes and Sulpice, 2009).

The recent advances in 'omics' technologies now make it feasible by integrating across different levels of genomic information, such as gene expression (transcriptomics), protein content (proteomics) and metabolite content (metabolomics), to better understand biomass production, improve crop breeding, and obtain ecological inference about the corresponding selective pressure acting on these QTLs (Fernie and Schauer, 2009; Keurentjes and Sulpice, 2009; Kliebenstein, 2009). Several studies have demonstrated the significance of metabolite QTLs (mQTLs) in enhancing our fundamental understanding of the genetic architecture regulating naturally variable phenotypes and the influence of this

basic research on agriculture and plant breeding (Keurentjes et al., 2006; Schauer et al., 2006; Schauer et al., 2008; Lisec et al., 2009).

Genomic approaches, by associating genome-scale analysis such as transcript profiling with a targeted phenotype such as measurements of specific metabolites have accelerated the study of the quantitative genetics that underlie phenotypic variation and can help in identifying uncharacterized networks or pathways (Rowe et al., 2008). The integration of metabolomics and genetics has been of great value in the study of plants and has provided powerful knowledge of the origin and maintenance of natural variation (Kliebenstein, 2009). These methods allow the quantification of highly diverse metabolites that can be associated with specific genetic markers, mRNA transcripts, and enzyme activities, allowing linkage between variation from genetics to phenotypes, such as plant growth (Koornneef et al., 2004).

The fundamental goal of metabolomics is to quantify the level of all intermediates of metabolism (Büscher et al., 2009). The recent progress in genome sequencing and high-throughput phenotyping technologies have facilitated increased coverage of all metabolites and have thus helped plant breeders to comprehensively elucidate the association between genetic and phenotypic variation (Phillips, 2008; Kliebenstein, 2009; Chan et al., 2010). On the basis of their physiological function and occurrence, metabolites can be categorized as primary or secondary, although in some cases it is difficult to make a clear distinction between primary and secondary metabolism as interactions between the two categories are inextricably present (Pichersky and Gang, 2000; Carrari et al., 2006). Biosynthetically, most secondary metabolites are derived from primary metabolic pathways and the genes of primary metabolism can serve as a pool from which similar genes of secondary metabolites can evolve. Primary metabolites are involved in basic physiological processes such as photosynthesis and respiration and there is an unequivocal relationship between primary metabolism and plant growth and development (Lisec et al., 2008). Changes in secondary metabolic networks such as those of the carotenoids or flavonoids can be achieved without any major pleiotropic effect concerning growth and development (Lorberth et al., 1998; Mann et al., 2000; Muir et al., 2001). In contrast, major perturbation of the networks of the primary metabolism such as sucrose biosynthesis or the tricarboxylic acid (TCA) cycle has a strong detrimental impact on plant performance related to whole-plant growth and development (Trethewey et al., 1998). Primary metabolites tend to be widespread, even universal, in occurrence and their functions are largely based upon empirical studies that reveal that they function in central carbon and nitrogen metabolism and include amino acids, sugars, fatty acids, and vitamins (Zangerl et al., 1997; Fritz et al., 2006; Price et al., 2008). Sugars such as glucose and sucrose and, recently, trehalose-6-phosphate have been demonstrated to act in the signalling of plant metabolic and sugar status and also to be involved in plant growth and development (Schluepmann et al., 2003; Gibson, 2005; Kolbe et al., 2005). Contrarily, secondary metabolites serve an ecological

function and are involved in defending plants against microbes and insects and are often connected to cell signalling and responses to biotic and abiotic stress (Zangerl et al., 1997; Wobus and Weber, 1999; Garg et al., 2002; Scheible et al., 2004)

Associations between plant and seed primary metabolites and plant growth and biomass, as well as seed quality, have been detected in a series of natural accessions, introgression and recombinant inbred populations (Schauer et al., 2006; Meyer et al., 2007; Prinzenberg et al., 2010; Skogerson et al., 2010; Toubiana et al., 2012). Although weak correlations are generally observed between growth and the levels of individual metabolites (Meyer et al., 2007), more significant relationships have been observed between biomass and a specific combination of metabolites (canonical correlation) (Lisec et al., 2009; Prinzenberg et al., 2010). Significant correlations between biomass, enzyme activities and metabolite content have been demonstrated in Arabidopsis (Cross et al., 2006). Further evidence for connectivity between plant growth and development and primary metabolism is derived from the coincidence of mQTLs with quantitative trait loci (QTLs) for whole-plant biomass and yield-related traits (Schauer et al., 2006; Lisec et al., 2007). Although recently Toubiana et al. (2012) have attempted to find links between tomato seed metabolites and plant phenotypes, and though extensive research exists on metabolic profiling for tomato fruit quality, no proper study exists which adequately covers the relationship between seed metabolite profiles, seed quality and seedling vigour (Schauer et al., 2006; Schauer et al., 2008).

The concept of genetical genomics in which traditional QTL analysis is integrated with gene expression and metabolic profiling has greatly improved our understanding of the genetic basis underlying complex traits (Jansen and Nap, 2001; Keurentjes et al., 2007) and provides researchers with additional tools to resolve metabolic, regulatory and developmental pathways (Kloosterman and Oortwijn, 2010). This is a useful methodology in studying molecular perturbations in biological systems. Several studies have used this approach, focusing on natural variation in different organisms and diverse types of populations (Schadt et al., 2003; Bystrykh et al., 2005; Keurentjes et al., 2006; Keurentjes et al., 2007)) and have detected extensive genetic regulation of gene expression. As molecular networks are also influenced by diverse environmental conditions, a comprehensive understanding of biological systems thus requires studying them across multiple environments. The proposed strategy by Li et al. (2008) using so-called generalized genetical genomics (GGG) for integration of multiple factors (genetics and sensibly chosen environments), is a crucial step towards understanding the environmental perturbation of molecular networks. In our present study, we used such a GGG approach for metabolic profiling using GC-TOF-MS on 100 recombinant inbred lines (RILs) of tomato to elucidate the genetic regulation of variation in the tomato seed metabolome. This GGG model may prove to be useful in the investigation of the mechanisms that contribute to complex variations in the tomato seed metabolome during germination and early seedling stage.

Major metabolic switches occur during the seed transition period from reserve accumulation to seed desiccation and germination, resulting in the accumulation of distinct metabolites. Efficient seedling establishment depends on the extent of accumulation of reserves during seed maturation and their subsequent efficient mobilization during seed germination (Fait et al., 2006). Based on the above considerations between the metabolic status of a plant system and growth and the established analytical power of metabolic profiling, in the present study we tried to unravel the link between seed metabolism and seed and seedling quality. For this purpose we used a tomato recombinant inbred line (RIL) population derived from a cross between *Solanum lycopersicum* and *S. pimpinellifolium*, which in previous studies showed a strong resolution for seed and seedling quality phenotypes (Kazmi et al., 2012; Khan et al., 2012). To unravel the metabolic switches affecting these traits, we quantified the metabolite content of tomato seeds at two developmental stages: dry and 6h-imbibed seeds.

Materials and Methods

Plant material

The tomato RIL population was obtained from a cross between *Solanum lycopersicum* cv. Moneymaker and *S. pimpinellifolium* CGN 15528 (Voorrips et al., 2000). The 100 RILs in this population were genotyped for a total of 865 Single Nucleotide Polymorphism (SNP) markers in F7 and seeds were harvested from F8 plants. 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;(Kazmi et al., 2012; Khan et al., 2012). The marker data were used to construct a genetic linkage map consisting of 12 individual linkage groups corresponding to the 12 chromosomes of tomato.

Growth conditions and seed collection

The RIL population was grown twice under controlled conditions in the greenhouse facilities at Wageningen University, the Netherlands. The population was grown under long day conditions (16h light and 8h dark) and the day and night temperatures were maintained at 25 and 15°C, respectively. The basic dose of fertilizer was uniformly supplied to all the RILs. Seeds were collected from healthy mature fruits and subsequently treated with 1% hydrochloric acid (HCl) for 1.5h to remove the pulp sticking onto the seeds. The solution of tomato seed extract with diluted HCl was passed through a fine mesh sieve and washed with tap water to remove pulp and HCl. The seeds were disinfected by soaking in a solution of trisodium phosphate ($Na_3PO_4.12H_2O$). 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). The cleaned seeds were dried for 3d 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.

Phenotyping of seed and seedling traits of the RIL population

Seed and seedlings of the same harvests were phenotyped in our previous study (Khan et al., 2012).

Extraction, derivatization and analysis of seed metabolites by GC-TOF-MS

We used an extraction method previously described by (Roessner et al., 2001) with some modifications. About 30mg of seeds per sample were homogenized using a micro dismembrator (Sartorius) in 2 ml tubes with 2 iron balls (2.5 mm) precooled in liquid nitrogen at 1500 rpm. For each sample 700µl methanol/chloroform (4:3) was added together with the standard (0.2mg/ml ribitol) and mixed thoroughly. After 10 minutes of sonication, 200µl Milli-Q water was added to the mixture followed by vortexing and centrifuging (5 min, 13500 rpm). The methanol phase was collected in a glass vial and 500µl of methanol/chloroform was added to the remaining organic phase and kept on ice for 10 minutes. Then 200µl Milli-Q was added followed by vortexing and centrifuging (5 min, 13500rpm). Again, the methanol phase was collected and mixed with the other collected phase. Subsequently, 100µl was dried overnight in a speedvac centrifuge at 35°C (Savant SPD121). The GC-TOF-MS method previously described by (Carreno-Quintero et al., 2012) was used with some minor modifications. Detector voltage was set at 1600V. Raw data was processed using the chromaTOF software 2.0 (Leco instruments) and further processed using the Metalign software (Lommen, 2009), to extract and align the mass signals. A signal-to-noise ratio of 2 was used. The output was further processed by the Metalign Output Transformer (METOT; Plant Research International, Wageningen) and the mass signals that were present in less than 3 RILs were discarded. Out of all the mass signals, centrotypes were formed using the MSclust program (Tikunov et al., 2012). 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 spectral 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 statistically analyzed by using the rank product method (Breitling et al., 2004) to identify differentially changed metabolites with the Bioconductor 'RankProd' package. The data were log2 transformed before statistical analysis. Significantly changed metabolites showed a false discovery rate (FDR) < 0.05. The FDR value in the rank product was obtained with 1000 random permutations. 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; ANOVA was also performed using R statistics (version 2.14.1, http://www.r-project.org/).

QTL Analysis

Data were 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 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 to 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: a LOD score of 3.0 was found to be significant (Breitling et al., 2008).

Integrated analysis of seed and seedling phenotypes and metabolite profiles

The relationship between seed and seedling biomass and metabolite profile was measured by simple Spearman correlation between the seed and seedling 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" (version 2.14.1, http://www.r-project.org).

Canonical correlation analysis (CCA)

Canonical correlation analysis measures 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 and seedling 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 (version 2.14.1, http://www.r-project.org). All procedures were applied after the missing value estimation followed by normalization of the metabolic matrix. Metabolites that are correlated at p value <0.02 with the phenotypic traits of seed and seedling were included in the canonical correlation analysis and were named as signature metabolites.

Results

Fluctuations in metabolites are associated with seed developmental stages

To determine the metabolic status in two different developmental stages in the seeds of a *Solanum lycopersicum* cv. Moneymaker ('MM') *x Solanum pimpinellifolium* ('Pimp') RIL population of 100 lines (Voorrips et al., 2000) and their parents, we used an in-house gas chromatography–time of flight–mass spectrometry (GC-TOF-MS) metabolomics platform. This GC-TOF-MS platform mostly identifies primary metabolites based on comparison with reference spectra (NIST and http://csbdb.mpimp-golm.mpg.de/csbdb/gmd/gmd.html). In the present study we could detect 160 metabolites in total and the chemical nature of 66 of these metabolites could be identified. These metabolites are commonly present in all living organisms and consisted of 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 biosynthetic pathways, such as glycerol-3-phosphate, ethanolamine, amino acids and their precursors, sugars, and some other common metabolic end products (Supplemental Table S5.1).

The parents of the RIL population revealed significant differences in their metabolite levels at both developmental stages (dry and 6h imbibed seeds). Metabolites showed significant differences in branched chain and aromatic amino acids, fatty acids, glutamate and ethanolamine, GABA, myo-inositol, phosphoric acid, sugars and organic acids. In general the level of metabolites was significantly higher in dry Pimp seeds as compared to the MM parent. The levels of 22 known metabolites including tricarboxylic acid cycle (TCA) and amino acid–associated metabolites were significantly higher in Pimp than in MM, whereas MM showed only higher levels of 4 metabolites, namely galactonic acid, urea, monomethyl phosphate and GABA in dry seeds (Figure 5.1A). However, metabolite levels increased significantly in MM after 6 hours of imbibition and in this case the levels of 22 metabolites were significantly higher in the MM parent as compared to Pimp with 6 metabolites at significantly higher levels (Figure 5.1B).

The data reveals that metabolite levels are subject to regulation by both genotype and environment. For example, levels of myo-inositol, serine and asparagine are higher in the Pimp parent under both environments whereas galactonic acid, monomethyl

phosphate, GABA and N-acetyl glutamic acid are higher in the MM parent in both environments (Figure 5.1A and B). In contrast, phosphoric acid is significantly higher in Pimp and urea in MM dry seeds but in 6h imbibed seeds both metabolites decreased differentially in the two parents to become significantly different in opposite directions. This is similar to several other metabolites for which the levels are strongly affected by the environment (Figure 5.1A and B). While comparing the metabolites of the same parent, it was observed that the levels of most metabolites significantly decreased in Pimp after 6 hours of imbibition whereas MM general displayes a significant increase (Figure 5.1C and D). The majority of the 160 metabolites were detected in both parents and in more than 90% of the RILs.

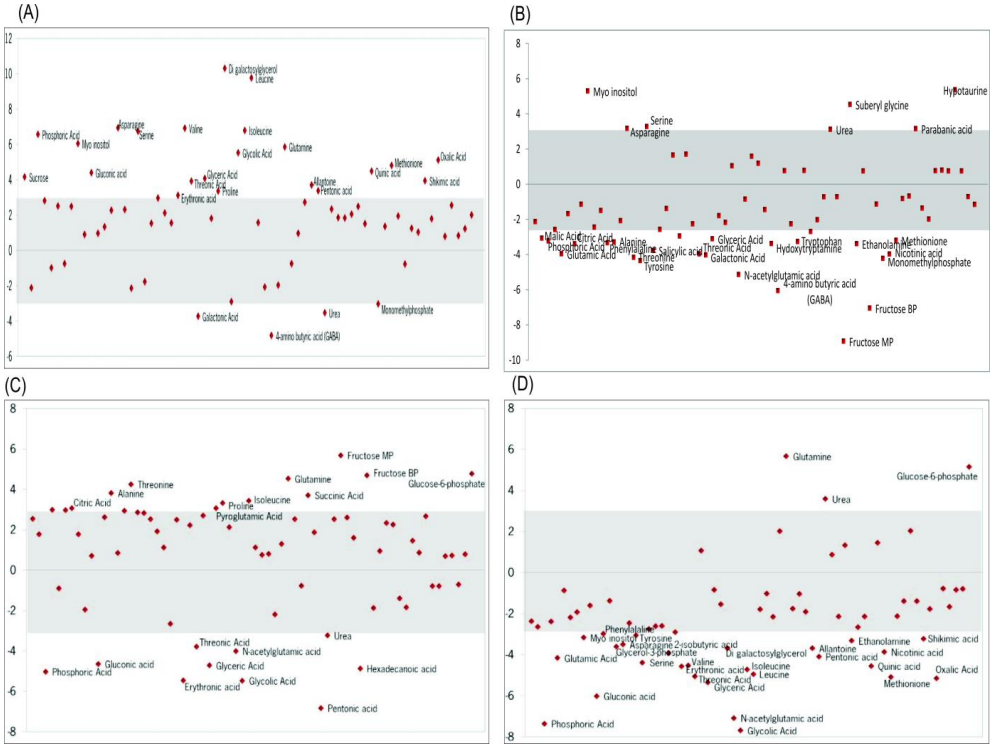


Figure 5.1. Significant differences between *S. lycopersicum* and *S. pimpinellifolium* based on dry (**A**) and 6h imbibed seeds (**B**). Metabolites above the grey box are significantly higher in *S. pimpinellifolium* while metabolites below the grey box are significantly higher in *S. lycopersicum*. (**C,D**) Differences of metabolite level between dry seeds and 6h imbibed seeds for *S. lycopersicum* (C) and *S. pimpinellifolium* (D) respectively. Metabolites above or below the grey box are respectively more or less abundant in 6h imbibed seeds. Metabolites are sorted on the x-axis from high abundant on the left and low abundant on the right. Metabolites indicated with a name are significantly different. p values are calculated with a student t-test (n=3).

For the analysis of the RIL population 50 lines were used for dry seeds and 50 for 6h imbibed seeds metabolomic profiling with smart selection of the lines, containing equal allele distributions from both parents at each selected locus (Li et al., 2009). Variations in the status of the major metabolites of the RIL population, at the two developmental stages, are presented in Figure 5.2. Quantitative changes in the levels of the several metabolites, including alcohols, sugars, organic acids and fatty acid compounds were associated with the change in seed developmental stage from dry to 6h imbibed (Figure 5.2). The levels of organic acids, such as, galactonate, glycolate, glycerate, erythronic acid, phosphoric acid, quinate and threonate, decreased intensely upon imbibition. The concentrations of amino acid and their precursors were invariant between dry and 6h imbibed seeds. Other compounds, such as alpha-hydroxybutyrate and the sugars xylofuranose, sucrose and the TCA-cycle intermediate oxalate also exhibited a significant decrease upon imbibition, whereas the other TCA-cycle metabolites declined even further upon imbibition. There was an increase in the concentrations of monomethyl phosphate, the organic acids parbanic acid and pentonic acid, and the TCA-cycle intermediate citrate in association with imbibition. In contrast, the levels of gluconate, quinate, shikimate and succinate were significantly reduced in the imbibed seed, whereas the levels of glycerine, aspartate, asparagine and hypotaurine increased significantly in the imbibed state. Similarly, the levels of most sugars declined but the levels of the sugar phosphates glucose-6-phosphate and glycerol-3-phosphate increased significantly.

Metabolites belonging to the same functional group are highly correlated across the population

We carried out a combined metabolite and phenotypic trait correlation analysis of dry and 6h imbibed seeds separately. For this purpose we created a correlation matrix of all pairwise comparisons among all 160 individual metabolites and 24 seed- and seedling traits by performing Spearman rank correlation analysis for all pairs of measured traits across the whole population. The seed and seedling traits included Fresh Root weight (FrRt), Dry Root weight (DrRt), Fresh Shoot weight (FrSh), Dry Shoot weight (DrSh), Hypocotyl Length (HypL), Seed Size (SS), Seed Length (SL), Seed Circularity (SC), Imbibed Seed Size (ImbSS), Imbibed Seed Length (ImbSL), Imbibed Seed Circularity (ImbSC), Main Root path Length (MRL), Total Root Size (TRS), Lateral Root number (LRn), Lateral Root density per Branch zone (LRd_Bz), Fresh Shoot weight Without nutrients (FrShWn), Dry Shoot weight Without nutrients (DrShWn), Fresh Root weight Without nutrients (FrRtWn) and Dry Root weight Without nutrients (DrRtWn). These seed and seedling traits were measured during our previous study (Khan et al 2012) and the seeds of the same batch were used for the metabolomics analysis. Spearman's rank correlation coefficients (Rs) and accompanying false discovery rate (FDR) corrected P values (P_{BH}; Benjamini-Hochberg) are provided in

Supplemental Table S5.2 (6h imbibed seeds) and 3 (dry seeds). The hierarchical clustering revealed several ''hot spots'' where most of the highly correlated phenotypic traits were clustered separately from metabolites (Supplemental Figure S5.1 and S5.2). It was notable that metabolites belonging to the same biochemical pathways or having similar functions clustered together in this matrix. For example, 10 of the 17 amino acids in dry seeds and 12 in the 6h imbibed seeds clustered together. Fourteen of these amino acids (82%) displayed absolute correlation coefficients ranging from Rs 0.51 to 0.95 (P_{BH}= 0 to 0.003) in both dry and 6h imbibed seeds. On the other hand organic acids clustered in several small groups, especially in 6h imbibed seeds on both sides of the phenotypic traits with several unknown metabolites clustering in between them. The clustering of unknown metabolites suggests a link with known metabolites. It was observed that metabolites having a significant positive correlation with phenotypic traits were clustered separately due to the strong absolute correlations within the metabolites (for example, 14 of the amino acids had significant positive correlations with more than 2 phenotypic traits). There were also metabolites that had weak correlations but clustered together with phenotypic traits, due to lack of correlation with any other metabolite (e.g. melezitose in dry seeds) or due to relatively a stronger correlation with the metabolite correlating with the phenotypic traits, such as glyceric acid, erythronic acid, glycolic acid in 6h imbibed seeds that clustered with phenotypic traits due to their strong correlation with galactonate. In general, the metabolites of 6h imbibed seeds produced higher numbers (429) of significant correlations (p value < 0.02) with the 24 seed and seedling traits, as compared to dry seeds (235 correlations). Further, the percent negative correlations of 6h imbibed seed metabolites with phenotypic traits was minimal (8%) compared with dry seeds (25%) (Supplemental Tables S5.2 and S5.3).

Canonical correlation reveals a closer association among seed, seeding biomass and a specific combination of metabolites as compared to single correlations

In a first step the predictive power of individual metabolites with respect to seed and seedling biomass was investigated by Spearman rank correlation analysis for all pairs of 160 measured metabolites and 24 phenotypic traits. It was observed that with the exception of few that correlated moderately, the majority of the metabolites correlated only weakly with phenotypic traits (Supplemental Tables S5.2 and S5.3). The highest absolute correlations were found between unknown metabolites (RI_1008, RI_1494, RI_997), which only correlated with dry and imbibed seed length (SL and ImbSL) and yielded an absolute Spearman R value of 0.58 each in the dry seed. The rest of the metabolites and phenotypic traits were weakly correlated in the dry seeds. The same trends were observed in 6h imbibed seed metabolites in terms of strength of correlation of individual metabolites. The highest absolute correlation was found for an unknown metabolite (RI_1644) with ImbSS

which yielded a Spearman R value of 0.602. Other metabolites of the imbibed seeds that correlated moderately with the phenotypic traits included both known (galactonate, glycerol-3-phosphate, quinate, valine) and unknown metabolites (RI_1008, RI_1135 RI_1494, RI_1858, RI_2023, RI_2196, RI_2210, RI_997). The rest of the metabolites were weakly correlated with phenotypic traits.

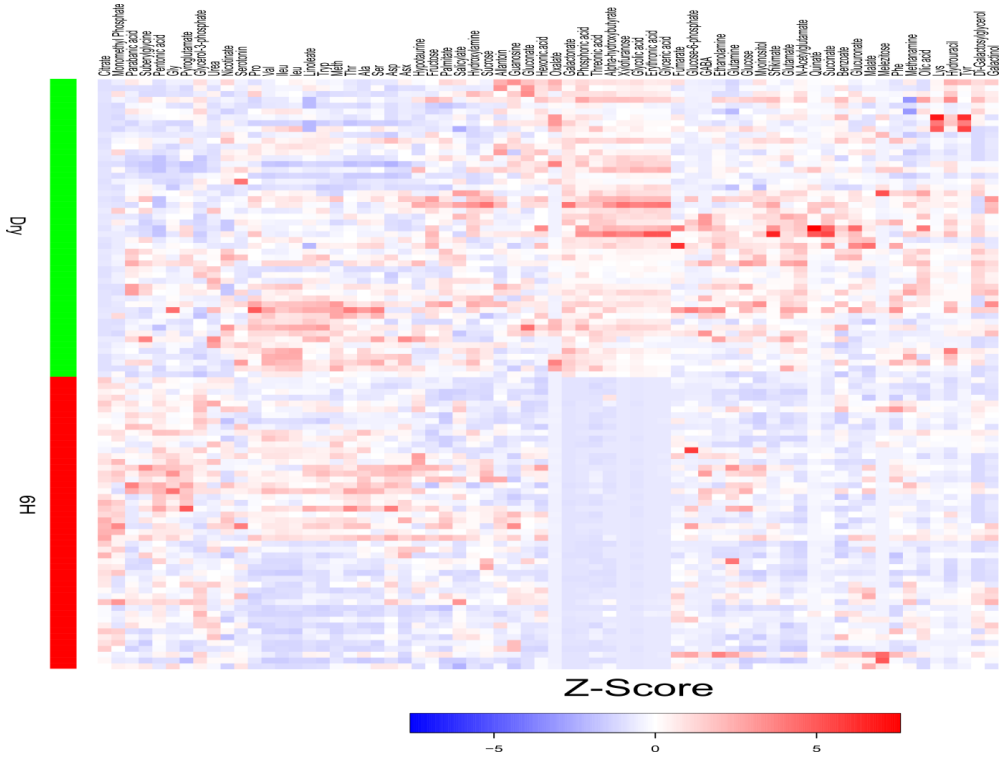


Figure 5.2. Metabolite profiles in dry and 6h imbibed seeds of tomato. Metabolite levels between dry and 6h seed developmental stages were compared. The vertical Green and Red bar at the left represent the RIL lines for which metabolite abundance was measured for dry and 6h imbibed seeds respectively. The relative abundance of each metabolite for each RIL line is shown. 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, from blue to red.

Therefore, we applied canonical correlation analysis (CCA) in a second step to predict the association between phenotypic traits and specific combinations of metabolites. Canonical correlation analysis is a multivariate technique usually used to assess associations between two separate data sets measured on the same experimental units (Gittins, 1985; Zwiers and Von Storch, 2004). Such canonical combination of metabolites has been used in previous analyses to predict certain phenotypic traits (Lisec et al., 2007; Meyer et al., 2007; Sulpice et al., 2010). We selected signature metabolites that displayed significant

correlation at p<0.02 with the phenotypic trait as the canonical variates. The top 10 (dry seeds) and top 34 (imbibed seeds) signature metabolites correlating with FrRt and top 8 and 30 signature metabolites correlating with FrSh are presented in Table 5.1 as examples, whereas details related to the canonical association among all phenotypic traits and specific combination of metabolites are provided in Figure 5.3 and Supplemental Table S5.4. The highest absolute correlations found with FrRt was by unknown metabolites (RI_2336) in dry seeds and RI_1494 in 6h imbibed seeds, which yielded values of -0.405 and +0.542, respectively.

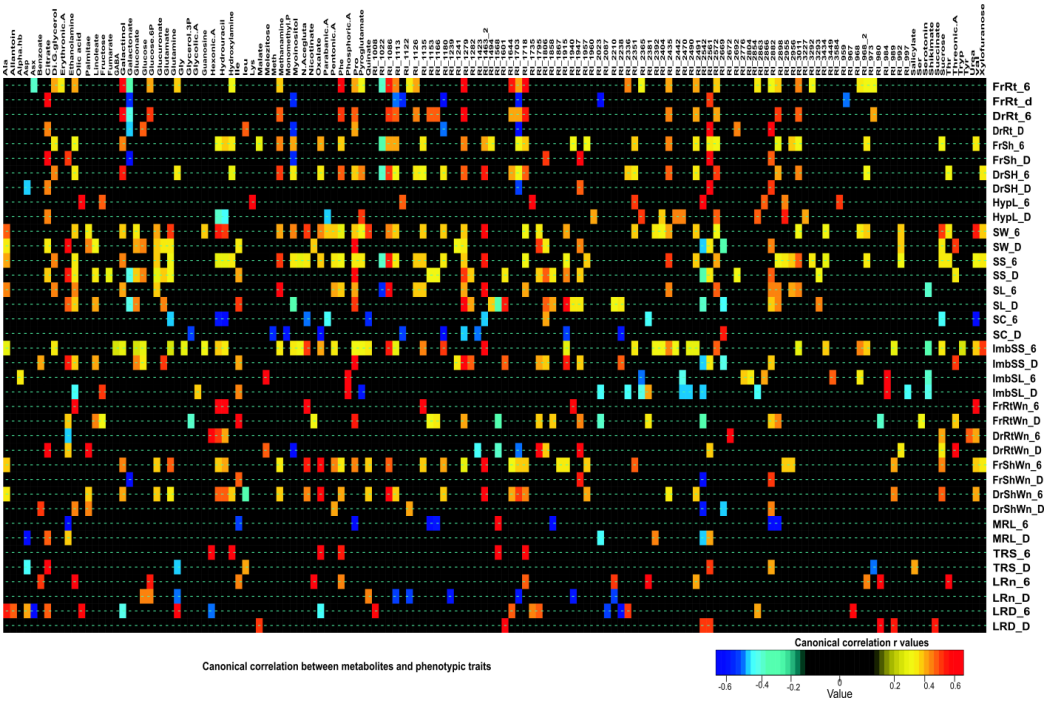


Figure 5.3. Heat map of canonical correlations between signature metabolites and phenotypic traits of seed and seedlings. The names of metabolites are given at the top of the heat map while the names of the phenotypic traits are given on the right side. The number of signature metabolites correlating at p value < 0.02 with phenotypic traits are represented in squares in the rows in front of each phenotypic trait.. Each square represents the Spearman correlation coefficient for canonical correlation. The yellow to red squares represent positive while the light green to blue represent negative correlations. The intensity of the color indicates the strength of the correlation. The color key is given at the bottom. Individual canonical correlation coefficients can be found in Table 5.1 and Supplemental Table S5.4. Phenotypic traits, FrRt (Fresh Root weight), DrRt (Dry Root weight), FrSh (Fresh Shoot weight), DrSh (Dry Shoot weight), HypL (Hypocotyl Length), SW (Seed Weight), SS (Seed Size), SL(Seed Length), ImbSS (Imbibed Seed Size), ImbSL (Imbibed Seed Length), SC (Seed Circularity), FrRtWn (Fresh Root weight under nutrientless condition), DrRtWn (Dry Root weight under nutrientless condition), FrShWn (Fresh Shoot weight under nutrientless condition), DrShWn (Dry Shoot weight under nutrientless conditions), MRL (Main Root path Length), TRS (Total Root Size), LRn (Lateral Root number) and LRD_Bz (Lateral Root Density per Branch zone). While _6 and _D means metabolites of 6h imbibed and dry seeds respectively.

Table 5.1 A. List of top 10 (dry) and 34 (imbibed) signature metabolites of 6-hour imbibed seeds ranked according to the strength of the canonical correlation with Fresh Root weight.

Dry Seed			6-hour imbibed seed		
Metabolites	Corr	PV	Metabolites	Corr	PV
RI_2336	-0.58738	1.14E-05	RI_1494	0.586836	9.35E-06
RI_1423	-0.58709	1.15E-05	RI_1858	0.586745	9.39E-06
Citrate	0.554084	4.39E-05	Galactonate	0.579253	1.30E-05
RI_1806	-0.53064	0.000105	Quinate	0.574681	1.57E-05
Oxalate	-0.52118	0.000146	RI_1795	0.519167	0.000132
Gluconate	-0.51012	0.000212	RI_1135	0.516018	0.000148
RI_1166	-0.50632	0.000241	RI_1153	0.453598	0.001063
RI_2692	0.495148	0.000346	RI_2392	0.452476	0.001098
RI_1153	-0.46093	0.000977	RI_3293	0.443731	0.001405
RI_968_2	-0.46071	0.000983	RI_1239	0.437952	0.001649
			RI_1022	0.436059	0.001736
			FA118.1	0.427944	0.00216
			RI_1806	0.424549	0.002363
			Glucuronate	0.424525	0.002364
			Di.Galactosylglycerol	0.42193	0.002531
			RI_2669	0.418684	0.002753
			RI_2491	0.417443	0.002843
			N.Acetylglutamate	0.405122	0.003882
			Benzoate	0.403259	0.004065
			Glycerol.3.phosphate	0.401659	0.004229
			RI_2776	0.398056	0.004618
			RI_3011	0.395867	0.004869
			Erythronic.acid	0.388755	0.005771
			RI_980	0.383449	0.006535
			RI_1180	0.382562	0.006671
			RI_1126	-0.36972	0.008933
			RI_2898	0.368704	0.009138
			Asx	-0.36779	0.009326
			Lys	0.365152	0.009885
			RI_989	0.358837	0.011343
			RI_2956	0.354939	0.012332
			RI_3449	0.354869	0.01235
			RI_1086	0.353062	0.012834
			RI_984	0.352342	0.013031

Table 5.1 B. List of top 8 (dry) and 30 (imbibed) signature metabolites ranked according to the strength of the canonical correlation with Fresh Shoot weight.

Dry seed metabolism			6-hour imbibed seed metabolism		
Metabolites	Corr	PV	Metabolites	Corr	PV
Citrate	0.727119	4.84E-09	Galactonate	0.457701	0.000944
Gluconate	-0.6335	1.35E-06	RI_2491	0.453741	0.001059
RI_2692	0.570352	2.31E-05	RI_1494	0.44004	0.001557
RI_2087	0.561323	3.32E-05	N.Acetylglutamate	0.438627	0.001618
Oxalate	-0.5052	0.00025	Oxalate	0.435911	0.001743
Ethanolamine	0.498347	0.000313	RI_1135	0.434297	0.001821
RI_1940	0.483789	0.000495	RI_1644	0.423791	0.00241
RI_2956	0.472626	0.000694	Phosphoric.acid	0.419481	0.002697
			RI_1153	0.413123	0.003175
			RI_3292	0.41016	0.003423
			RI_2404	0.40302	0.004089
			RI_1858	0.401172	0.00428
			Quinate	0.392777	0.005245
			leu	0.392677	0.005257
			RI_1241	0.38578	0.006189
			RI_2692	0.384025	0.006448
			Lys	0.382669	0.006654
			RI_1703	0.378481	0.007329
			RI_2669	0.369081	0.009062
			Di.Galactosylglycerol	0.368677	0.009143
			val	0.368537	0.009172
			RI_984	0.364934	0.009932
			FA118.1	0.363666	0.010212
			RI_3293	0.360195	0.011014
			Ileu	0.359635	0.011149
			RI_1806	0.3563	0.011978
			Thr	0.354453	0.01246
			RI_2210	0.350918	0.013429
			RI_2023	0.348504	0.014127
			RI_2776	0.348136	0.014236

Although the correlation is highly significant (p = 0.0042 and 5.57E-05), it can only explain 16 and 29% of the variance in dry and 6h imbibed seeds, respectively. Other significantly correlated metabolites with FrRt included citrate, oxalate, gluconate and 6 unknown metabolites in dry seeds (Table 5.1). Their individual contributions to the explained

variance ranged from 9% to 14%, whereas the individual contribution of other significantly correlating metabolites of imbibed seeds ranged from 11% to 28% (Supplemental Table S5.2). In contrast, CCA resulted in much stronger correlations of 0.69 and 0.925 for dry and imbibed seeds, respectively. These values corresponded to 48% and 85% of variance explained by the linear combination of metabolites, almost 1-4 times more than explained by any individual metabolite. In all the 18 phenotypic traits tested, the individual contribution ranged from 9 to 25% while the canonical correlation explained from 34% (MRL with 6h imbibed seeds) to a maximum of 95% (seed size with 6h imbibed seeds metabolites; Supplemental Table S5.4). The highest number of correlating metabolites was for seed traits, such as dry seed size and weight, and imbibed seed size. In general, higher canonical correlations were observed for 6h imbibed seeds metabolites and seed and seedling biomass, contrary to seedling RSA traits (MRL, TRS and LRn) where the canonical variate was higher for dry seeds (Figure 5.3 and Supplemental Table S5.4). These results suggest that variation in growth coincides with specific combinatorial changes of metabolites rather than fluctuation in only a few individual metabolites. The most remarkable results emerging from the CCA are (1) the number of signature metabolites, correlating significantly ($p<0.02$) with phenotypic traits for 6h imbibed seeds is 2-6 times greater for almost all of the phenotypic traits, as compared to dry seeds; (2) About 80-90% of the metabolites detected in 6h imbibed seeds are positively correlated with seedling traits, while 40-80% of the metabolites detected in dry seeds are negatively correlated with seedling traits (Figure 5.3, Table 5.1 and Supplemental Table S5.4). This suggests that even at early imbibition mobilization and breakdown of major seed reserves occurs and results in increased concentrations of those metabolites that are utilized for growth and development of the seedling (King and Gifford, 1997).

Identification of metabolic QTLs (mQTLs)

To inspect the genetic variation controlling tomato seed metabolism we used the generalized genetical genomics model (GGG**)** for the analysis of genetic variation of seed metabolites or the interaction between genetics and controlled environmental changes (Li et al., 2008). This generalized approach helped us to map quantitative trait loci (mQTLs) for both the genetic (G QTLs) as well as genetic-by-environment QTLs (G x E QTLs). We performed QTL mapping analysis by applying a linear model:

$$Y= G + GxE +E + \varepsilon$$

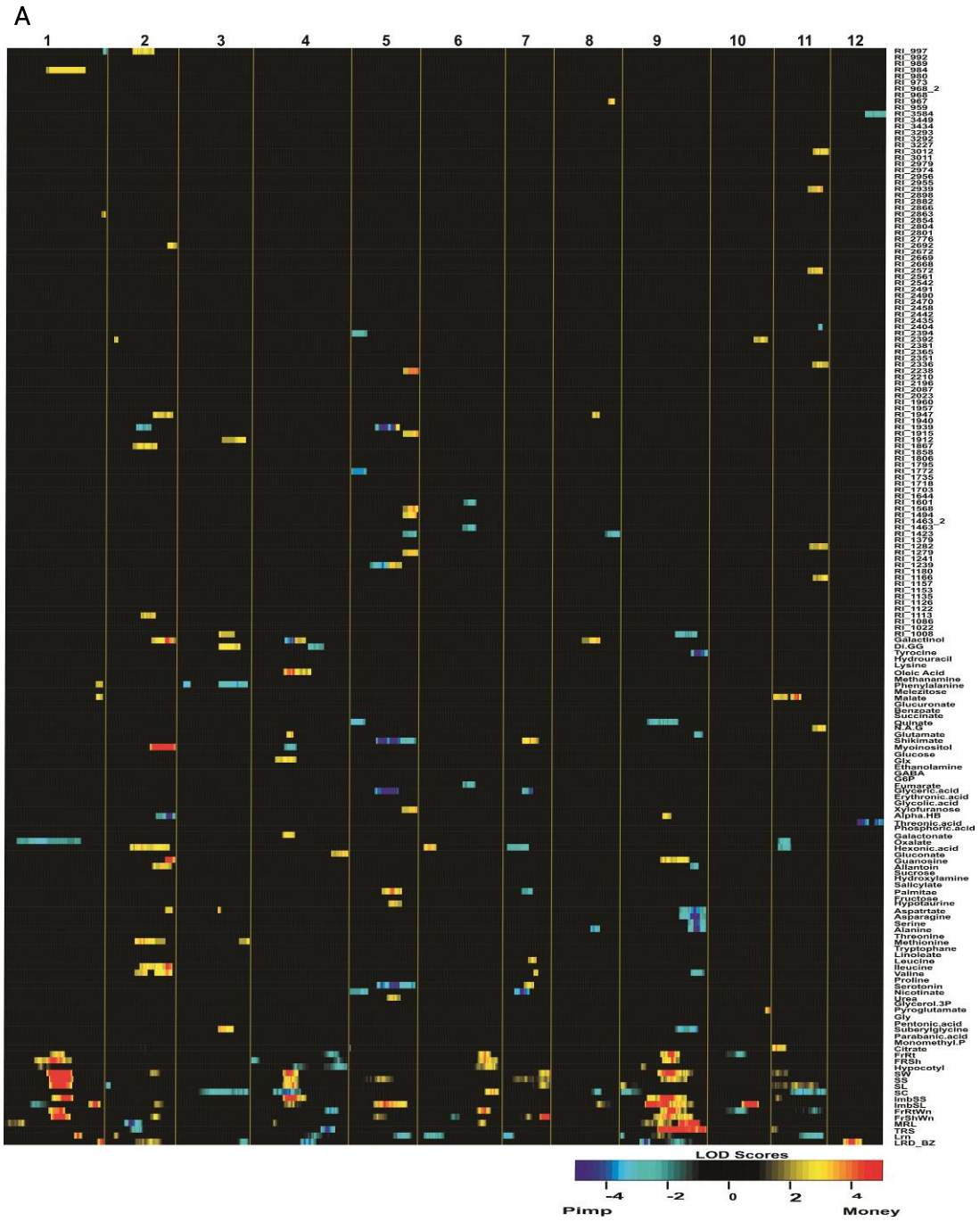
where Y is the matrix of measured metabolites, ε is the residual error and separated log10 probability values (called LOD scores hereafter) are generated for the environment (E), genetic (G) and genetic-by-environment (G x E) linkages. Significant thresholds of QTLs were determined by permutation analysis (n=1000, $p<0.01$) by randomizing the genotypes

over each metabolite and were strictly set to LOD >3 for all metabolites. By R/qtl analysis, using the linear model, we identified mQTLs for 60% known metabolites (72 QTLs), ranging from one to 5 QTLs per metabolite (Supplemental Table S5.5B, Figure 5.4A), and for 32% of the unknown metabolites (40 mQTLs). We detected 34 G x E QTLs (Supplemental Table S5.5B and Figure 5.4B).

An overview of co-locating QTLs between metabolites and between phenotypic traits

Traits are declared co-locating when the 1-LOD intervals of the traits are overlapping with each other (Supplemental Table S5.5A and S5.5B, Figure 5.4A and 5.4B). As content of metabolites with co-locating QTLs are expected to correlate, and with an increasing number of shared QTLs the correlation may increase, a higher incidence of co-location of mQTLs than the observed was predicted, as many of the metabolites were strongly correlated. Several small and large groups of co-locating mQTLs were observed on different chromosomes with the most dominant clusters of 14, 5, 10, 8, 9 and 7 mQTLs on chromosomes 2, 4, 5, 7, 9 and 11 respectively (Figure 5.4). The cluster of 14 mQTLs on chr 2 includes 9 known and 5 unknown metabolites. All of these metabolites have the same genetic effect from the MM parent. The known metabolites included 4 amino acids (valine, Ieucine, methionine and aspartate), allantoin, guanosine, hexonic acid and the sugar alcohols myoinositol and galactinol.

These metabolites are also strongly correlated to each other. For example, the absolute correlation value between valine and Ieucine is 0.96 and they also significantly correlated with other co-locating metabolites (myoinsitol and RI_1113). Similarly, myo-inositol was significantly correlating with galactinol and other metabolites in either both or only dry or imbibed seed metabolite profiles (Supplemental Figure S5.1 and S5.2 and supplemental Tables S5.2 and S5.3). Further, we also found metabolites with co-locating QTLs in this cluster but the levels of these metabolites were not correlating in either dry or imbibed seeds. For example, galactinol and hexonic acid had QTLs at the same position and with the same genetic effect as the amino acid group but both of these metabolites had no correlation with this amino acid group in this cluster of co-locating QTLs. It was also observed that 11 of the mQTLs (valine, leucine, methionine, allantoin, hexonic acid, myo-inositol, galactinol, RI_1113, RI_1867, RI_1947 and RI_997) at this locus were also overlapping with 4 phenotypic QTLs (LRD_Bz, FrShWn, ImbSL, and SW). The concentration of these metabolites also correlated significantly with at least one or more phenotypic traits (Supplemental Figure S5.1 and S5.2 and supplemental Tables S5.2 and S5.3). Similarly, a group of 3 amino acids (galactonate, glycine, glutamate), oleic acid and galactinol were co-locating with each other on chromosome 4 as well as with seed traits (SW, SS, SL, ImbSS and ImbSL; Fig 5.4A, Supplemental Table S5.5). In this case only glutamate and glycine were strongly correlating with each other and there was no significant correlation among the other metabolites. Yet, all of these metabolites correlated significantly with seed and


A
1
2
3
4
5
6
7
8
9
10
11
12
RI_997
RI_992
RI_989
RI_984
RI_980
RI_973
RI_968_2
RI_968
RI_967
RI_959
RI_3584
RI_3449
RI_3434
RI_3293
RI_3292
RI_3227
RI_3012
RI_3011
RI_2979
RI_2974
RI_2956
RI_2955
RI_2939
RI_2898
RI_2882
RI_2866
RI_2863
RI_2854
RI_2804
RI_2801
RI_2776
RI_2692
RI_2672
RI_2669
RI_2668
RI_2572
RI_2561
RI_2542
RI_2491
RI_2490
RI_2470
RI_2458
RI_2442
RI_2435
RI_2404
RI_2394
RI_2392
RI_2381
RI_2365
RI_2351
RI_2336
RI_2238
RI_2210
RI_2196
RI_2087
RI_2023
RI_1960
RI_1957
RI_1947
RI_1940
RI_1939
RI_1915
RI_1912
RI_1867
RI_1858
RI_1806
RI_1795
RI_1772
RI_1735
RI_1718
RI_1703
RI_1644
RI_1601
RI_1568
RI_1494
RI_1463_2
RI_1463
RI_1423
RI_1379
RI_1282
RI_1279
RI_1241
RI_1239
RI_1180
RI_1166
RI_1157
RI_1153
RI_1135
RI_1126
RI_1122
RI_1113
RI_1086
RI_1022
RI_1008
Galactinol
DI.GG
Tyrocine
Hydrouracil
Lysine
Oleic Acid
Methanamine
Phenylalanine
Melezitose
Malate
Glucuronate
Benzoate
Succinate
Quinate
Ala
Glutamate
Shikimate
Myoinositol
Glucose
Gly
Ethanolamine
GABA
G6P
Fumarate
Glyceric.acid
Erythronic.acid
Glycolic.acid
Xylofuranose
Alpha.HB
Threonic.acid
Phosphoric.acid
Galactonate
Oxalate
Hexonic.acid
Gluconate
Guanosine
Allantoin
Sucrose
Hydroxylamine
Salicylate
Palmitae
Fructose
Hypotaurine
Aspatrtate
Asparagine
Serine
Alanine
Threonine
Methionine
Tryptophane
Linolenate
Leucine
Ileucine
Valine
Proline
Serotonin
Nicotinate
Urea
Glycerol.3P
Pyroglutamate
Gly
Pentonic.acid
Suberylglycine
Parabanic.acid
Monomethyl.P
Citrate
FrRt
FRSh
Hypocotyl
sW
SS
SL
SC
ImbSS
ImbSL
FrRtWn
FrShWn
MRL
TRS
Lrn
LRD_8Z
LOD Scores
-4
-2
0
2
4
Pimp
Money

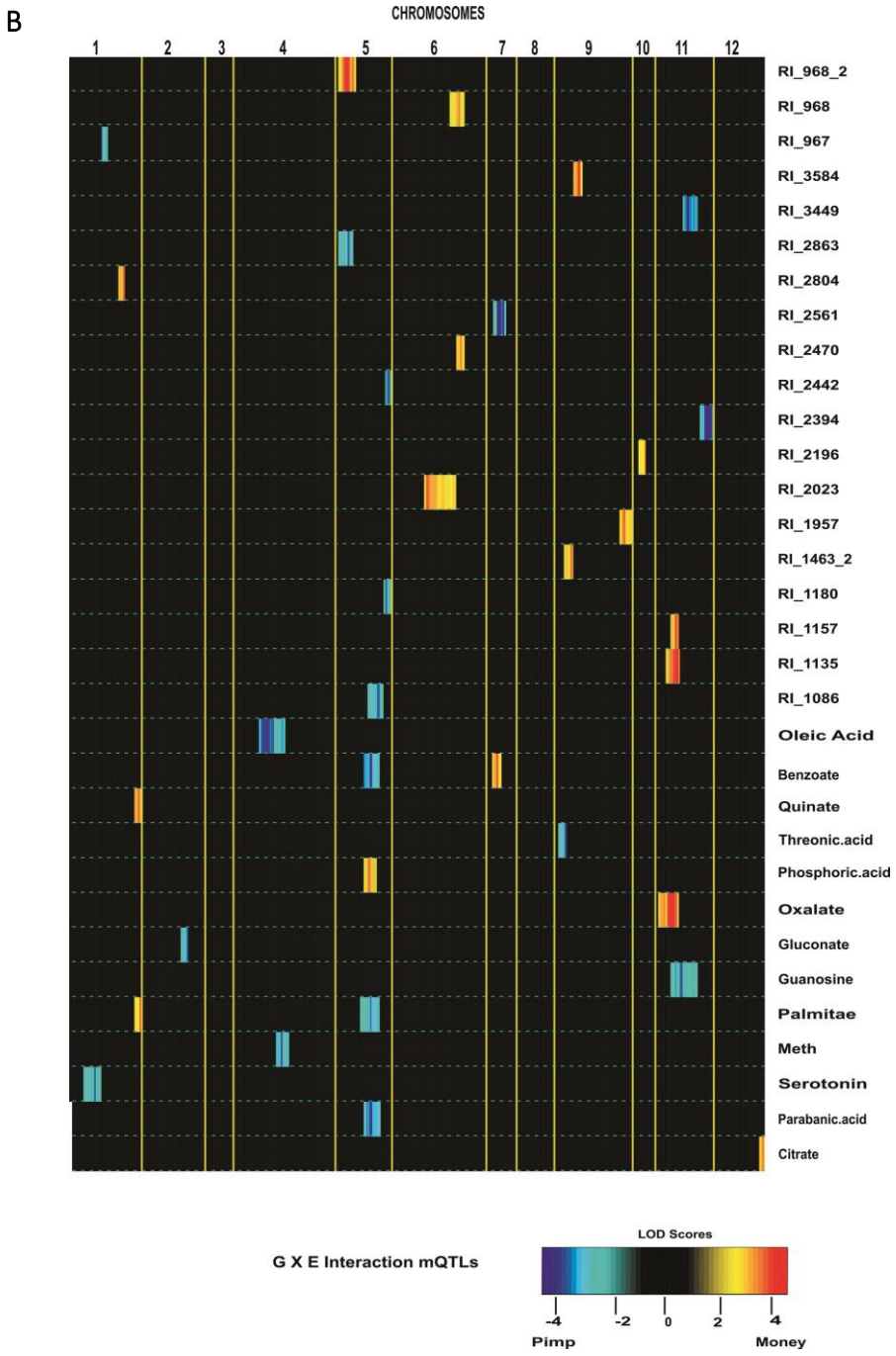


Figure 5.4. (A) Genomic locations of genetic mQTLs identified for metabolite accumulation showing overlapping with phenotypic QTLs previously identified Chapter 2. (B) Genomic locations of G x E mQTLs identified for metabolite accumulation.

Tomato chromosomes are shown by numbers (1-12), with centi-Morgans ascending from left to right; chromosomes are separated by yellow lines. Coloured cells indicate significant QTLs. 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. While for the phenotypic traits the significant threshold was previously set to LOD >2.5 by permutations analysis (N=1000, p<0.01). The LOD colour 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 content. QTL positions, LOD scores and effects are provided in Supplemental Table S5.5A (genetic mQTL), Supplemental Table S5.5B (GxE mQTLs) and Supplemental Table S5.6 (seed and seedling phenotypic QTLs). The phenotypic traits include FrRt= Fresh Root weight, FrSh= Fresh Shoot weight, HypL= Hypocotyl length, SW= Seed Weight, SS= Seed Size, SL= Seed Length, SC= Seed Circularity, ImbSS= Imbibed Seed Size, ImbSL= Imbibed Seed Length, MRL= Main Root Length, TRS= Total Root Size, LRn= Lateral Root number and LRD_Bz= Lateral Root Density per Basal zone. Some of the metabolites are also abbreviated: N.A.G= N-acetylglutamate, Di.GG= Di-Galactosylglycerol, Alpha.HB= Alpha-Hydroxyisobutyric acid and G6P= Glucose-6-Phosphate.

seedling traits, especially in 6h imbibed seeds (Supplemental Figure S5.1 and S5.2 and supplemental Tables S5.2 and S5.3).

In general, we observed three options in term of correlation and co-location: (1) there is co-location of QTLs between metabolites or between metabolites and phenotypic traits, but there is no correlation among them. For example in the middle of chr 5, three metabolic QTLs (urea, hypotaurine and palmitic acid) are occurring in the same interval and also co-locate with three phenotypic traits (FrShWn, ImbSL and SS) but there is no significant correlation between the metabolites or between the metabolites and phenotypic traits; (2) there is both co-location and correlation: for example 4 amino acids (alanine, serine, asparagine and aspartic acid) have a cluster of mQTLs on chromosome 9 that also coincide with the QTLs for subryglycine, valine, allantoin, glutamate, and tyrosine and the content values of these metabolites are mostly positively correlated; (3) there is no co-location but still correlation: for example the metabolic QTLs on chr 9 are not co-locating with the major phQTL detected on this chromosome but most of these metabolites are positively correlating with seed and seedling traits.

Discussion

In the current study we investigated the genetic basis of natural variability in seed primary metabolism and its response to a changing environment. To achieve this goal, we employed metabolite profiling of seeds from a tomato RIL population of *S. lycopersicum* and *S. pimpinellifolium*. Subsequently, we conducted an integrated analysis together with data from previous phenotypic studies of tomato seed and seedling quality to try to predict the phenotypes as a function of metabolite composition (Khan et al., 2012). The integrated heterogeneous data of metabolite profiles and phenotypic traits were investigated by simple and canonical correlation, as well as QTL analysis, which allowed us to comprehensively understand the complex nature of seed metabolism. A number of

correlations and identification of putative mQTLs coinciding with phQTLs helped us to predict the link between seed metabolic status and seedling establishment.

The transition from dry to the 6h imbibed seed stage is associated with a distinct metabolic switch

Profiling seed metabolites in diverse environments revealed the complex nature of seed metabolism in a changing environment. Growth and metabolic activities are arrested in the dry seed; however, upon imbibition the quiescent seed resumes normal and high level of metabolic activity (Bewley et al., 2013). We opted 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 although no genetic differences in developmental stage are expected yet. . Metabolite concentrations were differentially regulated by the genotype as well as the environment. Major metabolic switches were observed between dry and 6h imbibed seeds. In the dry state 33% of the known metabolites including most amino acid, sugars and organic acids were significantly higher in the Pimp parent compared to the MM parent where only 6% were significantly higher (Figure 5.1A). In the 6h imbibed seeds the situation was reversed with 36.3% of the metabolites significantly higher in MM compared to Pimp. On the basis of this transition, metabolites were either (1) genetically (G) regulated, irrespective of the environment, such as myoinositol, serine and asparagine, which are significantly higher in Pimp in both dry and imbibed seeds and galactonic acid, GABA and monomethylphosphate in MM (Figure 5.1A and B), (2): symmetrically environmentally regulated (E), such as most organic acids that were concomitantly down regulated and glutamine and glucose-6-phosphate that were up regulated in both parents in imbibed seeds (Figure 5.1C and D) or (3) differentially environmentally regulated in different genotypes (G X E), such as urea which was significantly higher in MM in the dry state and steeply declined in imbibed seeds, while at the same rate it increased in imbibed Pimp seeds (Figure 5.1C and D). Such fluctuations in the RIL population have been presented in Figure 5.2 where the genetically regulated metabolites have variable colours in the heat map in both the dry and imbibed states. Metabolites such as oxalate, galactonate, phosphoric acid, threonic acid and citric acid appear to be more regulated by environment than genotype and they could clearly be distinguished in the dry and 6h imbibed seeds. Though the level of most metabolites significantly decreased in the Pimp parent upon imbibition, the level of most metabolites, especially amino acids, were significantly up-regulated in the MM parent.

This increase and decrease in metabolite levels upon imbibition in both parents is in accordance with the presence of seed food reserves. The period of reserve accumulation in seeds is associated with a major reduction of primary metabolites including amino acids,

sugars, organic acids and polyols, suggesting their utilization and incorporation into storage reserves such as protein, starch, and fatty acids (Fait et al., 2006). Contrary to maturation, the transition period from reserve accumulation to seed desiccation is associated with a major metabolic switch due to degradation of these seed reserves, resulting in the accumulation of distinct sugars, organic acids, nitrogen-rich amino acids, and shikimate-derived metabolites. The level of these metabolites in the seed is dependent on the rate of reserve accumulation during the seed maturation period. The breakdown of storage proteins in loblolly pine (*Pinus taeda L)* was correlated with a substantial increase in the free amino acid pool in the seedling (King and Gifford, 1997). Thus the higher concentration of major metabolites in Pimp dry seeds indicates an early desiccation and degradation of the seed reserve. This is also in accordance with the fast germination capability of the Pimp parent in our previous study (Kazmi et al., 2012). The reduction in metabolites following imbibition implies that primary metabolites are rapidly consumed to support the metabolic switch toward enhancing biosynthetic processes required for early germination. Although Bewley (1997), and Eastmond and Graham (2001) have documented that mobilization of oil and protein reserves in oil seeds occurs following radical protrusion, Fait et al. (2006) have shown that in Arabidopsis active metabolic processes are already initiated during seed imbibition and that significant reduction occurred in the level of most metabolites that had accumulated during maturation drying. On the other hand, the MM parent is a relatively late-germinating genotype under normal conditions. This late germination might be due to a low rate of reserve food mobilization during maturation drying and the rapid increase in the level of major metabolites following imbibition is perhaps the start of degradation of major reserves during this period. Thus the switch from a dry seed to 6h imbibed seeds is associated with release of energy from degradation and remobilization of reserve food for seed germination and subsequent seedling growth.

Genetic and environmental regulation and correlation of metabolic traits

One of the objectives of our study was to analyze the RIL population over different environments or developmental stages to identify metabolic fluxes in the changing environment, in addition to the effect of genetic heritability. The QTL analysis revealed that metabolic abundance was regulated by both genotype and environment. For example, citrate content significantly increased in 6h imbibed seeds and we detected a significant QTL for the genetic x environmental effect, but could not detect any QTL for this metabolite for the genetic component. Benzoate, oxalate and phosphoric acid strongly decreased during imbibition and we identified QTLs for each of these metabolites for the genetic x environmental effect but, again, no genetic QTLs could be detected. Further, metabolites such as serotonine, palmitic acid, gluconate and quinate are examples of metabolites that are differentially regulated both by genotype, environment and their interaction which resulted in both G and G x E QTLs. On the other hand, metabolites such as myoinositol,

serine and galactinol are examples of metabolites that are only strongly regulated by genetic factors, which is confirmed by strong genetic QTLs with significant LOD scores. As a whole, we identified a significantly higher number of genetic mQTLs (112) as compared to 34 genetic x environmental mQTLs. This suggests that higher percent of the metabolic signature of the RIL population is regulated by the genetic components rather than environmental.

In general, the possibility to find mQTLs in the RIL population was low as mQTLs were identified for only 46% of the 160 metabolites (112 mQTLs in total), as compared to the phenotypic seed and seedling traits for which we detected 115 phQTLs for only 19 traits studied with a minimum of 3 to a maximum of 9 QTLs per trait. However, the identification of only a low percentage of mQTLs may be due to the fact that in many cases many independent pathways may be responsible for the synthesis of the same metabolite and therefor single genetic loci might have minimal effect on the metabolite levels. For example, we could not find any QTL for the amino acid glycine and it has been documented that glycine is the product of two very short pathways, one using threonine and the other serine as precursors (Levine and Hwa, 2007). Thus, these pathways have multiple controllers and any one genetic locus may not significantly change the level of metabolites and, therefore, cannot be detected as mQTL (Ferrara et al., 2008). Nevertheless, we observed that strongly correlated metabolites mapped to identical positions, suggesting that these are regulated by a common genetic factor. Although co-locating QTLs can be the result of independent closely linked genetic factors, such coinciding QTLs are expected to occur more or less randomly by chance. Here we provide evidence of co-regulation of biologically-related pathways. When plotted against their genomic positions, seven of such suggestive QTL clusters can be seen on 6 different chromosomes, several of which also coincide with QTLs for seed and seedling traits. There is strong correlation between levels of metabolites present in these QTLs hotspots, and many of them are also correlating with seed and seedling biomass. Metabolites providing the most remarkable evidence of functional clustering are the amino acids that group together in both the correlation matrix and the QTL hotspot on chromosome 9. Another group of amino acids including valine, isoleucine, methionine and aspartate, exclusively mapped to chromosome 2, together with other known metabolites such as allantoine, guanosine, hexonic acid, myo-inositol, galactinol and 4 unknown metabolites, whereas valine also clustered with leucine, serotonine and shikimate on chromosome 7.

Contents of co-locating QTLs are expected to correlate and the correlation may become stronger with increasing numbers of co-locating QTLs (Lisec et al., 2007). Such correlation may be caused by common genetic factors, e.g. regulatory or biosynthesis genes. However, there are cases of co-locations of mQTLs or phQTLs and no correlation in the corresponding content of metabolites or phenotypic traits, and vice versa. For example, guanosine is co-locating with several seed and seedling trait QTLs on chromosome 9 but

the concentration of this metabolites in both dry and 6h imbibed seeds is not correlating with any seed or seedling trait. Such situation could be due to strong interactions of the co-locating QTL with other loci or the metabolite may be under different additional metabolic and environmental controls. Similar findings have been reported where correlations between metabolites and phenotypic traits were not simply due to coincidence of QTLs for these metabolic and phenotypic traits (Lisec et al., 2007; Meyer et al., 2007).

The correlation between levels of amino acids and other metabolites and their precursors is in accordance with our understanding of metabolic physiology. The success of germination and subsequent seedling vigour is associated with the amount, degradation and mobilization of food reserves (Fait et al., 2006; Penfield et al., 2006; Bewley et al., 2012). The efficiency and mobilization of reserves during seed germination apparently depends on the amount of reserve accumulation during seed maturation (Fait et al., 2006). Our data suggests that correlation-based clustering of metabolites could be used as a biomarker for changes in the flux of certain metabolic pathways. However, our data also reveals that these metabolites are under the control of different genetic regulators, resulting in a distinctive genetic mapping, even within a group of strongly correlating metabolites, which hints at the complex nature of metabolic regulation.

Canonical integration of metabolites with seed and seedling quality phenotypic traits

Our results show that pairwise correlation analysis of seed and seedling phenotypes and single metabolites resulted in weak correlations which could only explain a maximum of 9-25% of the total variance observed in the phenotypes. Such correlations have been found previously between single metabolite- and morphological traits in tomato (Toubiana et al., 2012). These correlations of single metabolites with phenotypic traits are, however, substantially greater than those observed in other studies, e.g., only 7% when comparing single metabolites with seedling biomass in Arabidopsis (Meyer et al., 2007). In contrast, canonical correlation analysis yielded much stronger correlations, in the present study ranging from a minimum of 60% to a maximum of 98% between different phenotypic traits and specific combinations of metabolites. The CCA revealed that a combination of the levels of a larger number of metabolites rather than a few individual metabolites, points at a close correlation with seed mass and, indirectly, seedling growth. Thus, the present study supports previous findings of (Meyer et al., 2007; Lisec et al., 2008; Sulpice et al., 2010.These results demonstrate that variation in seed mass and seedling growth may coincide with characteristic combinatorial fluctuations of metabolite levels, whereas individual metabolites may change largely independently without any major changes in growth.

Inspection of metabolites highly ranked in CCA provided strongly linked clusters in which both known metabolites of central metabolic pathways, as well as unknown metabolites, are strongly correlated with morphological traits. Of the unknown metabolites

detected here, 26 of the dry seeds and 38 of the 6h imbibed seeds were strongly represented with a minimum of 3 to a maximum of 15 per phenotypic traits. Among the known metabolites hexose sugars and hexose phosphates such as glucose, sucrose, fructose, glucose-6-phosphate and glycerol-3-phosphate were highly represented by correlating with more than one phenotypic trait. These sugars play an important role by linking carbon flow from photosynthesis and starch metabolism with cell wall formation. A high glucose level probably maintains 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). Sucrose is also the major transport form of carbon from source to sink tissue and thus represents the border between carbohydrate synthesis and its utilization at the whole plant level (Meyer et al., 2007). In addition, the oxidative pentose phosphate pathway also provides substrates for nucleic acid synthesis and for lignin, polyphenol and amino acid synthesis, as well as glycolysis. Another highly ranked sugar phosphate is glycerol-3-phosphate which is also known to play a major role in membrane/phospholipid biosynthesis. Metabolites such as ethanolamine and digalactosyldiacylglycerol (DGDG) are also highly ranked in the CCA with seed and seedling traits. Ethanolamine is an metabolite that plays an important role in membrane/phospholipid biosynthesis (Meyer et al., 2007). Increased levels of this metabolite resulted in significantly enhanced seedling growth in Arabidopsis (Kelly et al., 2003). DGDG is a major lipid in the photosynthetic membranes of oxygenic photosynthetic organisms (Awai et al., 2007) and is required for normal growth of Arabidopsis thaliana. A *dgd1* mutant in Arabidopsis, defective in a plant-type DGDG synthase, showed severe dwarfism with correlated decreases in DGDG content, chlorophyll content, light-harvesting complex II (LHCII) stability and photosynthetic capability (Dörmann et al., 1995).

However, the major groups that were strongly represented in the CCA were the organic and amino acids. Among the organic acids, citrate, gluconate, malate, oxalate, galactonate, pentonic acid, phosphoric acid, quinate, threonate and salicylate were present in the top-ranking metabolites with several seed and seedling traits in either dry or 6h imbibed seeds. Organic acid metabolism is of central importance at the cellular level for several biochemical pathways, including energy production, formation of precursors for amino-acid biosynthesis and at the whole organism level in modulating adaptation to the environment (Carrari et al., 2003). Organic acids also play a role as key components in the mechanisms that some plants use to cope with nutrient deficiencies, metal tolerance and plant-microbe interactions, operating at the root-soil interface. For example, citrate, malate and oxalate are some of the organic acids that are released from the roots of aluminium-tolerant plants and can form sufficiently strong complexes with Al^{3+} to protect plant roots (Ma et al., 2001). Among the amino acids, 16 were represented in the signature metabolites but Trp, Ser, Phe and Gln, in the dry seeds and Val, Lys, Leu, Thr, Ile, Ala, Glu, Gly, Pro and Tyr in 6h imbibed seeds, were the top ranking metabolites in the CCA. In

addition, the amino acid precursors N-acetylglutamate and pyroglutamate are also highly-ranked metabolites in the CCA. The most remarkable observation is that 15 of the amino acids are positively correlated with more than one phenotypic trait and only Asp is negatively correlated in dry seeds although it is also positively correlated in 6 h imbibed seeds. Our data show that 90% of the known metabolites detected here are positively correlated with seed and seedling traits in both dry and 6h imbibed seeds, although the number of metabolites and number of correlations is higher in 6h imbibed seeds. However, a higher percentage of the unknown metabolites is negatively correlated in dry seeds and, vice versa, in 6h imbibed seeds. The increase in correlations in the imbibed stage suggests that major reserve compounds of the dry seed are degraded during imbibition and are converted to those metabolites that are needed for energy supply for germination and subsequent seedling growth.

These observations support the findings of Toubiana et al. (2012) who showed that most seed metabolites displayed positive correlations with seed traits in tomato although they found a high degree of negative correlation between seed metabolites and morphological traits, such as fruit size , mature plant weight and harvest index. The strong representation of amino acids in the CCA hints at the significance of amino acid metabolic pathways in seedling vigour and establishment. Amino acids such as Gln, Glu, and Asp, which serve as central metabolites in nitrogen assimilation in plants are found amongst the most important metabolites. In addition, Ala, Ile, Leu, Met, Ser, Phe, Pro and Val, ranking highly in the CCA, are functional amino acids which are related to stress responses.

The positively correlating metabolites may play a major role as nutrients for plant growth and in defense against biotic and abiotic stress and it is logical that higher levels of these metabolites are accompanied by better growth of plants. There is growing evidence that, besides their role as building blocks of proteins and polypeptides, many amino acids regulate key metabolic pathways that are necessary for plant growth, maintenance and reproduction. The negatively correlating metabolites indicate an opposite pattern of change of distinct morphological traits, with respect to the specific quantities of metabolites across the population, or it indicates that high growth rates cause a depletion of those metabolites. Thus, a link between the metabolites ranking high in the CCA and seed and seedling quality phenotypes could be established as part of central metabolism which appears to be of key importance to seed quality and subsequent seedling vigour. It is also obvious from both the higher number as well as the higher explained variance of signature metabolites correlating with seed weight followed by major seedling traits such as root and shoot weight in the canonical correlation analysis that seed size is the major regulator of seed metabolism.

The probability of predicting biomass on the basis of the metabolic signature of seeds provides an opportunity for the use of metabolite profiles as biomarkers with high predictive power and could potentially revolutionize breeding for the improvement of seed

and seedling quality. Profiling the primary metabolome over different developmental stages or environments in a mapping population may result in generating new testable hypotheses and may improve our basic understanding of the genetic and environmental effects on the behavior of seed metabolites. 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 RIL population, 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 and seedling quality traits and forms a crucial step towards understanding different influences of genetic and developmental responses in tomato seeds.

Supporting information

Supporting information can be downloaded from
http://www.wageningenseedlab.nl/thesis/nkhan/SI/chapter5

Supplemental Table S5.1. List of known detected metabolites and their distribution among the two parents ('Pimp'= *S. pimpinellifolium* and 'MM'= *S. lycopersicum*) of the studied RIL population. Sheet 'metabolites_categories' in this table represents sorting of metabolites into categories and the CHEBI-IDs (http://www.ebi.ac.uk/chebi/init.do) of the uncharged molecule are given as unique identifiers.

Supplemental Table S5.2. Spearman Rs values, associated p Values and FDR (False Discovery Rate) corrected p values (Benjamini and Hochberg, 1995) for all pairwise correlations for all 160 metabolites of 6h imbibed seeds and seed/seedling phenotypes.

Supplemental Table S5.3. Spearman Rs values, associated p Values and FDR (False Discovery Rate) corrected p values (Benjamini and Hochberg, 1995) for all pairwise correlations for all 160 metabolites of dry seeds and seed/seedling phenotypes.

Supplemental Table S5.4. 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 with the phenotypic traits. The 6h imbibed and dry seed metabolites are shown in separate sub-tables one after the other respectively for each trait. CCcor (canonical correlation), CCPV (Canonical correlation p values), Pcor (Pearson correlation r values) and PPV (Pearson correlation p values).

Supplemental Table S5.5A and B. Genetic mQTLs for both known and unknown metabolites (Supplemental Table 5A) and GxE mQTLs for both known and unknown metabolites (Supplemental Table 5B)

1. Names of the metabolites.
2. Chromosomes on which the QTLs were detected.
3. 1-LOD support interval in centiMorgan.
4. The position of the QTL (it's highest peak).
5. LOD score (LOD score of 3.1 or above was calculated to be significant for mQTLs in this population).

Supplemental Table S5.6. Overview of significant QTLs associated with seed and seedling traits of the *S. lycopersicum* x *S. pimpinellifolium* RIL population (100 Lines).

1. Traits: 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, ImbSC= Imbibed Seed Circularity, FrShWn= Fresh Shoot weight under nutrientless conditions, FrRtWn= Fresh Root weight under nutrientless conditions, , DrShWn= Dry Shoot weight under nutrientless conditions, DrRtWn= Dry Root weight under nutrientless conditions, MRL= Main Root Length, TRS=Total Root Size, LRn= Lateral Root number, LRD_Bz= Lateral Roots Density per Branched zone.
2. Chromosomes on which the QTLs were detected.
3. 1-LOD support interval in centi-Morgan.
4. Nearest marker to the position of ìdentified QTL.
5. LOD score (LOD score of 2 or above was calculated to be significant for this population)
6. Additive effect; 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.
7. Percentage of variation explained by each QTL.
8. Percentage of total variation explained by genetic factors for a single trait as estimated by MapQTL.
9. Broad-sense heritability estimate for each trait.

Supplemental Figure S5.1. Heat map of correlations among all 160 metabolites of 6h imbibed seeds of 50 RILs with seed/seedling phenotypes. The names of phenotypic traits are coloured in green, known metabolites in blue and unknown metabolites in red. Each square represents the Pearson correlation coefficient between the seed metabolites and seed and seedling phenotypes of the column with that of the row. The 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). Individual correlation

coefficients can be found in Supplemental Table S5.2. Phenotypic traits, FrRt= Fresh Root weight, DrRt= Dry Root weight, FrSh= Fresh Shoot weight, DrSh= Dry Shoot weight, HypL= Hypocotyl Length=, SW= Seed Weight=, SS= Seed Size, SL= Seed Length, ImbSS= Imbibed Seed Size, ImbSL= Imbibed Seed Length, SC= Seed Circularity, FrRtWn= Fresh Root weight under nutrientless condition, DrRtWn= Dry Root weight under nutrientless condition, FrShWn= Fresh Shoot weight under nutrientless condition, DrShWn= Dry Shoot weight under nutrientless conditions, MRL= Main Root Length, TRS= Total Root Size, LRn= Lateral Root number and LRD_Bz= Lateral Root Density per Branch zone. SW2, SS2, SL2 and SC2 are seed weight, size, length, and seed circularity respectively of a second independent harvest of the RILs.

Supplemental Figure S5.2. Heat map of correlations among all 160 metabolites of dry seeds of 50 RILs with seed/seedling phenotypes. The rest of the characteristics of this figure are the same as described in the legend for Supplemental Figure S5.1.

References

Awai K, Watanabe H, Benning C, Nishida I (2007) Digalactosyldiacylglycerol is required for better photosynthetic growth of Synechocystis sp. PCC6803 under phosphate limitation. Plant and Cell Physiology **48:** 1517-1523

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

Bewley JD, Bradford KJ, Hilhorst HWM, Nonogaki H (2012) Seeds: Physiology of Development, Germination and Dormancy. Springer Verlag, New York

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

Büscher 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

Bystrykh L, Weersing E, Dontje B, Sutton S, Pletcher MT, Wiltshire T, Su AI, Vellenga E, Wang J, Manly KF (2005) Uncovering regulatory pathways that affect hematopoietic stem cell function using'genetical genomics'. Nature genetics **37:** 225-232

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:** 1380-1396

Carrari F, Urbanczyk-Wochniak E, Willmitzer L, Fernie AR (2003) Engineering central metabolism in crop species:: learning the system. Metabolic Engineering **5:** 191-200

Carreno-Quintero N, Acharjee A, Maliepaard C, Bachem CWB, Mumm R, Bouwmeester H, Visser RGF, Keurentjes JJB (2012) Untargeted metabolic quantitative trait loci analyses reveal a

relationship between primary metabolism and potato tuber quality. Plant Physiology **158:** 1306-1318

Chan EKF, Rowe HC, Hansen BG, Kliebenstein DJ (2010) The complex genetic architecture of the metabolome. PLoS Genetics **6:** e1001198

Cross JM, Von Korff M, Altmann T, Bartzetko L, Sulpice R, Gibon Y, Palacios N, Stitt M (2006) Variation of enzyme activities and metabolite levels in 24 Arabidopsis accessions growing in carbon-limited conditions. Plant Physiology **142:** 1574-1588

Dörmann P, Hoffmann-Benning S, Balbo I, Benning C (1995) Isolation and characterization of an Arabidopsis mutant deficient in the thylakoid lipid digalactosyl diacylglycerol. Plant Cell **7:** 1801-1810

Eastmond PJ, Graham IA (2001) Re-examining the role of the glyoxylate cycle in oilseeds. Trends in Plant Science **6:** 72-78

El-Lithy ME, Clerkx EJM, Ruys GJ, Koornneef M, Vreugdenhil D (2004) Quantitative trait locus analysis of growth-related traits in a new Arabidopsis recombinant inbred population. Plant Physiology **135:** 444-458

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

Fernie AR, Schauer N (2009) Metabolomics-assisted breeding: a viable option for crop improvement? Trends in Genetics **25:** 39-48

Ferrara CT, Wang P, Neto EC, Stevens RD, Bain JR, Wenner BR, Ilkayeva OR, Keller MP, Blasiole DA, Kendziorski C (2008) Genetic networks of liver metabolism revealed by integration of metabolic and transcriptional profiling. PLoS Genetics **4:** e1000034

Fritz C, Palacios-Rojas N, Feil R, Stitt M (2006) Regulation of secondary metabolism by the carbon–nitrogen status in tobacco: nitrate inhibits large sectors of phenylpropanoid metabolism. Plant Journal **46:** 533-548

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

Gibson SI (2005) Control of plant development and gene expression by sugar signaling. Current Opinion in Plant Biology **8:** 93-102

Gittins R (1985) Canonical analysis: a review with applications in ecology, Vol 88. Springer-Verlag Berlin

Jacobsson L, Park H, Wahlberg P, Fredriksson R, Perez-Enciso M, Siegel PB, Andersson L (2005) Many QTLs with minor additive effects are associated with a large difference in growth between two selection lines in chickens. Genetical Research **86:** 115-125

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

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

Kelly AA, Froehlich JE, Dörmann P (2003) Disruption of the two digalactosyldiacylglycerol synthase genes DGD1 and DGD2 in Arabidopsis reveals the existence of an additional enzyme of galactolipid synthesis. Plant Cell **15:** 2694-2706

Keurentjes JJB, Fu J, De Vos CHR, Lommen A, Hall RD, Bino RJ, van der Plas LHW, Jansen RC, Vreugdenhil D, Koornneef M (2006) The genetics of plant metabolism. Nature genetics **38:** 842-849

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 of the United States of America **104:** 1708-1713

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

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

King JE, Gifford DJ (1997) Amino acid utilization in seeds of loblolly pine during germination and early seedling growth (I. arginine and arginase activity). Plant Physiology **113:** 1125-1135

Kliebenstein DJ (2009) Advancing genetic theory and application by metabolic quantitative trait loci analysis. Plant Cell **21:** 1637-1646

Kloosterman B, Oortwijn M (2010) From QTL to candidate gene: Genetical genomics of simple and complex traits in potato using a pooling strategy. BMC genomics **11:** 158

Kolbe A, Tiessen A, Schluepmann H, Paul M, Ulrich S, Geigenberger P (2005) Trehalose 6-phosphate regulates starch synthesis via posttranslational redox activation of ADP-glucose pyrophosphorylase. Proceedings of the National Academy of Sciences of the United States of America **102:** 11118-11123

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

Levine E, Hwa T (2007) Stochastic fluctuations in metabolic pathways. Proceedings of the National Academy of Sciences of the United States of America **104:** 9224-9229

Li SB, Zhang ZH, Hu Y, Li CY, Jiang X, Mao T, Li YS, Zhu YG (2006) Genetic dissection of developmental behavior of crop growth rate and its relationships with yield and yield related traits in rice. Plant Science **170:** 911-917

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

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

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

Liu BH (1997) Statistical genomics: linkage, mapping, and QTL analysis. CRC

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

Lorberth R, Ritte G, Willmitzer L, Kossmann J (1998) Inhibition of a starch-granule–bound protein leads to modified starch and repression of cold sweetening. Nature Biotechnology **16:** 473-477

Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits.

Ma JF, Ryan PR, Delhaize E (2001) Aluminium tolerance in plants and the complexing role of organic acids. Trends in Plant Science **6:** 273-278

Mann V, Harker M, Pecker I, Hirschberg J (2000) Metabolic engineering of astaxanthin production in tobacco flowers. Nature Biotechnology **18:** 888-892

Meyer RC, Steinfath M, Lisec J, Becher M, Witucka-Wall H, Törjék O, Fiehn O, Eckardt Ä, Willmitzer L, Selbig J (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

Muir SR, Collins GJ, Robinson S, Hughes S, Bovy A, De Vos CHR, van Tunen AJ, Verhoeyen ME (2001) Overexpression of petunia chalcone isomerase in tomato results in fruit containing increased levels of flavonols. Nature Biotechnology **19:** 470-474

Penfield S, Li Y, Gilday AD, Graham S, Graham IA (2006) Arabidopsis ABA INSENSITIVE4 regulates lipid mobilization in the embryo and reveals repression of seed germination by the endosperm. Plant Cell **18:** 1887-1899

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

Pichersky E, Gang DR (2000) Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. Trends in Plant Science **5:** 439-445

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:** 1441-1461

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

Rocha JL, Eisen EJ, Dale Van Vleck L, Pomp D (2004) A large-sample QTL study in mice: I. Growth. Mammalian Genome **15:** 83-99

Roessner U, Wagner C, Kopka J, Trethewey RN, Willmitzer L (2001) Simultaneous analysis of metabolites in potato tuber by gas chromatography–mass spectrometry. 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

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

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

Scheible WR, Morcuende R, Czechowski T, Fritz C, Osuna D, Palacios-Rojas N, Schindelasch D, Thimm O, Udvardi MK, Stitt M (2004) Genome-wide reprogramming of primary and secondary metabolism, protein synthesis, cellular growth processes, and the regulatory infrastructure of Arabidopsis in response to nitrogen. Plant Physiology **136:** 2483-2499

Schluepmann H, Pellny T, Van Dijken A, Smeekens S, Paul M (2003) Trehalose 6-phosphate is indispensable for carbohydrate utilization and growth in *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences of the United States of America **100:** 6849-6854

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

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 E-T, 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. Plant Cell **22:** 2872-2893

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

Tonsor S, Alonso-Blanco C, Koornneef M (2004) Gene function beyond the single trait: natural variation, gene effects, and evolutionary ecology in *Arabidopsis thaliana*. Plant, Cell & Environment **28:** 2-20

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. Plant Journal **15:** 109-118

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

Wobus U, Weber H (1999) Sugars

Zangerl AR, Green ES, Lampman RL, Berenbaum MR (1997) Phenological changes in primary and secondary chemistry of reproductive parts in wild parsnip. Phytochemistry **44:** 825-831

Zwiers FW, Von Storch H (2004) On the role of statistics in climate research. International Journal of Climatology **24:** 665-680

Chapter 6

Using Heterogeneous Inbred Families (HIFs) to Confirm Natural Allelic Variation for Complex Seed and Seedling Phenotypes on Tomato Chromosomes 6 and 9

Khan N, Willems LAJ, Ligterink W, Hilhorst HWM

Abstract

Being a complex trait, the genetic basis of seed quality is poorly understood. Numerous tomato seed weight, seed size and seedling vigour related quantitative trait loci (QTLs) were previously identified in our study on the *Solanum lycopersicum (*cv. Moneymaker, 'MM') x *Solanum pimpinellifolium* (G1.1554, 'Pimp') tomato recombinant inbred line (RIL) population. Fine-mapping is usually a follow-up step for a detailed mapping and comprehensive characterisation of individual loci. However, prior to fine-mapping, confirmation and validation of the identified QTLs is essential and is generally applied in a practical breeding program. Here we report on the successful confirmation of some of those complex loci for seed and seedling traits on chromosomes 6 and 9 by analyzing near-isogenic lines (NILs). The NILs used for the QTL confirmation were so-called heterogeneous inbred families (HIFs) developed from the RIL population. These HIFs segregate for genomic regions spanning the QTLs under study.. The application of this procedure is described for QTLs related to a number of seed and seedling quality traits in tomato. Based on residual heterozygosity in the RIL population, candidate HIFs were screened with two CAPS markers for the identification of HIFs that were segregating for QTLs for seed size, seed weight and several seedling traits. Two segregating families were identified for each marker from linkage groups 6 and 9, respectively. The progeny of these HIFs was tested for the segregation of seed and seedling phenotypes for markers flanking different QTLs on these loci. NILs derived from each HIF had significantly different phenotypic values, confirming a number of loci on chromosome 6 and 9 that influence seed and seedling quality traits. One seedling quality trait QTL on chromosome 6 and eight seed quality trait QTLs on

chromosome 9 were confirmed. Subsequent fine-mapping and cloning of the causal genes for these QTLs could lead to the identification of novel genes controlling different associated genetic and physiological seed processes.

Introduction

Exploring the phenotypic variation between individuals, either of the same or of different species can improve our understanding of how species adapt to their environment, and to what extent are individual differences due to environmental effects, genetic effects, or both (Slate, 2005). However, to know whether one or many genes explain a large proportion of phenotypic variation or whether individual genes explain variation in several traits (pleiotropy), or how gene action depends on the environment and what are the evolutionary forces that maintain genetic variation, is a great challenge (Barton and Turelli, 1989). In the late 1980s, the advent of molecular markers and genetic maps has made it possible to map the genes that explained continuous variation. In the beginning most of the mapping studies were focused on humans and agriculturally important crops, but soon afterwards evolutionary biologists initiated quantitative trait locus (QTL) studies, especially in classical model organisms such as *Drosophila melanogaster* (Shrimpton and Robertson, 1988; Mackay, 1995).

In recent years, several new genomics resources and tools have become available that will greatly facilitate mapping quantitative trait locus (QTL) and cloning of the corresponding genes governing the phenotype in different genotypes. Tens of thousands of molecular markers, genome sequences, microarrays, and knock-out collections are being used for QTL mapping, thus facilitating the use of natural variation for gene discovery (Borevitz and Chory, 2004). However, despite substantial progress in cloning QTL genes, and even reducing some of them to Quantitative Trait Nucleotides (QTNs), QTL mapping and cloning remain a formidable task (Rikke and Johnson, 1998; Wu and Lin, 2006; Holland, 2007; Ron and Weller, 2007). Therefore deciphering the genetic basis of natural variation in quantitative traits faces a challenge because the variation is often continuous and because there are often extensive genotype x environment (GxE) interactions. In addition, as QTL mapping results in large genetic intervals and QTLs with large effects can be fragmented into multiple QTLs, explaining only a small proportion of the total variance also makes it a difficult task to determine the causal genes (Balasubramanian et al., 2009). Similarly, the phenotypic influence of individual QTLs is further complicated by the phenotypic variability resulting from segregation of other loci affecting the same trait (Loudet et al., 2005).

To identify the exact position and characterize the QTL additional methods and experiments are essential. Near-isogenic lines (NILs) that differ only for markers flanking a specific QTL could be one of the solutions to reduce the map position of a QTL (Robertson

et al., 1988; Kaeppler et al., 1993; Kooke et al., 2012). Investigation of the NILs allows the splitting of the large intervals of QTLs into smaller regions when they differ in smaller regions of the genome than identified by the QTL analysis (Tuinstra et al., 1997). Thus NILs are the perfect starting material for the fine-mapping and cloning of QTLs and can be used not only to map QTLs to smaller genomic intervals but also for phenotypic confirmation of QTLs. Depending on the availability of polymorphisms, in some cases the QTL can be mapped directly to the gene (Fridman et al., 2000; Kroymann et al., 2001). Thus, NILs present a common genetic resource in which direct comparison of two lines can be applied to estimate the phenotype conditioned by a QTL (Touzet et al., 1995; Tanksley and Nelson, 1996; Eduardo et al., 2005).

However, in spite of their importance for narrowing down QTL intervals, the use of NILs has been limited in crop breeding, mainly due to the considerable efforts required to develop suitable genetic material. NILs can be constructed through a variety of methods depending on the available resources, such as advanced backcrosses (BCs), recombinant inbred lines (RILs), doubled haploids (DHs), heterogenous inbred families (HIFs), or other mapping populations (Kooke et al., 2012). In all instances, however, the starting point is a cross between two genotypes which segregate in subsequent generations and, in most cases, one to several rounds of backcrossing and/or selfing are required to ultimately recover the desired genomic composition. However, from an advanced BC population, NILs carrying small introgressions from the donor parent, can be easily isolated by marker-assisted selection (MAS) (Kooke et al., 2012). Although initially developed from heterogeneous progeny of selected crosses, NILs preferably are homozygous. The genetic make-up is then fixed in so called “immortal” lines which can be used permanently and in many replications in numerous experiments. NILs can then be used to acquire better understanding of the magnitudes of QTL × Environment (QTL X E), QTL × Genetic (QTL x G) and QTL × QTL interactions (Eshed and Zamir, 1995, 1996), fine-map QTLs, eliminate undesirable effects caused by linkage drag and, eventually, perform positional cloning of the genes causal for the QTLs (Alpert and Tanksley, 1996). Thus NILs can serve many functions, ranging from help in breeding purposes to genetic dissection of complex quantitative traits.

HIFs are a set of lines derived from RILs which have residual heterozygosity spanning the locus of interest but are homozygous elsewhere (Loudet et al., 2005). Each RIL is then selfed and genotyped so that each homozygous genotype at the locus of interest can be identified and studied further. HIFs are not to be compared with the reference parental genotype but with one another within the descendants of the chosen RIL. Since HIFs originate from one RIL of the population, the genetic background of HIFs is a mixture of the two parents of the RIL population, unlike ‘conventional’ NILs. Molecular markers can be used to screen a population of HIFs derived from different inbreds to identify families that segregate for a specific genomic region in which a QTL has been detected for a

particular trait (Tuinstra et al., 1997). This procedure can be implemented to develop a series of NILs that contrast for a specific genomic region harbouring the QTL for a specific trait. This HIF concept is feasible and more effective as compared to a 'conventional' NIL strategy because there is no need to first create the NILs, which typically is laborious and requires several generations of backcrossing and marker-assisted selection.

Quantitative genetics has gained general acceptance as a tool for describing genetic variation of phenotypic characters in natural populations and for predicting their response to selection. Numerous quantitative trait analysis studies in different agriculturally important species have been carried out which have yielded a long list of genomic regions governing a wide range of important phenotypic traits (Schlichting and Pigliucci, 1995). Many of these QTLs have been cloned (Sweeney and McCouch, 2007) and the causal genes for several of these QTLs have been identified (Clouse and Sasse, 1998; Ligterink et al., 2012). The recent additions to the identification of various seed and seedling quality QTLs in tomato and Arabidopsis have enhanced 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).

We have previously analyzed a RIL population generated from *Solanum lycopersicum (*cv. Moneymaker, 'MM') and *Solanum pimpinellifolium* (G1.1554, 'Pimp') to identify QTLs for seed and seedling quality phenotypes (Khan et al., 2012) and seed germination phenotypes (Kazmi et al., 2012). Clusters of QTLs were observed across the 12 tomato chromosomes that influence seed dimensions, seedling biomass and seed germination in both control- and stress conditions. Many of these QTL hotspots were co-locating for seed and seedling traits, as well as for seed germination characteristics across different environments. These clusters of QTLs may be very useful for breeding purposes and can be used as molecular markers for improving seed quality. In our tomato RIL population, we identified HIFs on chromosome 6 and 9 that were segregating in the genomic regions in which QTLs were identified for seed and/or seedling phenotypes. The purpose of this study was to evaluate and confirm seed and seedling quality QTLs on chromosomes 6 and 9 by testing the HIFs that contrast at the QTL regions controlling seed and seedling phenotypes. Taking advantage of the residual heterozygosity and the large size of the RIL population, we were able to confirm QTLs for seed weight, size and length on chromosome 9 and hypocotyl length on chromosome 6.

Materials and Methods

Identification of QTLs

For identification of seed and seedling QTLs the *S. lycopersicum* (cv. Moneymaker, 'MM') × *S. pimpinellifolium* (G1.1554, 'Pimp') RIL population was tested under both nutrient starvation and normal nutrient conditions (Khan et al., 2012). The mapping software

MapQTL®6.0 (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®6.0.

Development, growth conditions and seed collection of heterogeneous inbred families

HIFs were derived from specific residual heterozygosity remaining in some of the F_8 RILs at the loci of interest on chromosomes 6 and 9. For each of these HIF lines, 20 plants were individually genotyped at the segregating markers. One HIF (HIF233A) segregating for a seedling QTL on chromosome 6 and four HIFs (HIF233B, 239, 241 and 259) ranging from 54.142 to 115.399 cM, segregating for seed and seedling traits on chromosome 9 were identified (Table 6.1). A set of NILs was selected from each of the segregating HIFs for chromosomes 6 and 9. The HIF233A (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 chromosome 9, the HIF233B set of NILs consisted of 3 lines with Pimp, 6 lines with MM and 1 line with a heterozygous background. Similarly, in HIF239 we could select 3 lines with Pimp, 2 lines with MM and 9 lines with a heterozygous background. In case of HIF241 we could select a set of 15 NILs of which 6 lines have a Pimp background, 4 lines an MM background and 5 lines a heterozygous background, whereas for HIF259 we could select 9 lines with Pimp, 4 lines with MM and 4 lines with a heterozygous background.

These lines were genotyped with CAPS markers across the tomato genome to determine the average heterogeneity of each HIF. These HIFs, along with the two parents, 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 until use.

Genotyping of HIFs

DNA was extracted using a previously described method (Cheung et al., 1993). PCR was performed on this DNA in a Bio-Rad S1000™ Thermocycler. The PCR reactions were conducted for 4 min at 95°C, followed by 35 cycles of a 20 sec denaturation step at 95°C, primer annealing; 30 sec at 55°C and 1 min at 72 °C followed by a 10 min extension step at 72°C. PCR products were used for restriction analysis with the desired enzyme. Restriction products were run on a 1.5% agarose gel stained with Gel Red and the genotype of the plant was assessed. The forward and reverse sequences of the primers used for the PCRs,

the physical position of the markers (in bp) and the enzymes used for the marker analysis are given in Table 6.1.

Table 6.1. HIFs, markers, physical positions and start and end points of the HIFs

HIFs		HIF233A	HIF239	HIF241	HIF259	HIF233B
Markers Used		STW553/554	STW1000/1001	STW1054/1055	STW1058/1059	STW1070/1071
bp(physical)		42710926	60987757	63638559	65492233	66097878
cM		92.0	65.002	82.424	98.225	105.399
cM Hetero	Low	76.151	54.933	78.009	94.46	98.748
	High	96.821	70.853	92.585	112.29	115.208
bp Hetero	Low	40009629	59830549	63211245	65062496	65914934
	High	43761285	62496837	64960323	66561219	67115755
Primers used	Forw.	TCTCACTTCC-CTACATTCTC	ACAAGAGGAG-CTGGATAC	GTTGGGAGG-TTTTTGAATTG	ATAAAAGAG-AGGTCGGGG	GTGAGTTTG-ATTTTGCACC
	Rev.	ATACCCATA-GACTTGCTG	AGGGGCAAA-GGGAGAAAA	GTACTTGGT-CGGGAAATG	GAAAAGGAGT-GATATCAAGGG	AATAACCTG-CTGTGGAGA
Enzyme		SalI	Indel marker	HindIII	HindIII	NheI

Phenotypic characterization

Phenotyping of seedlings

To characterize differences in seedling quality phenotypes, experiments were performed on the 20 NILs derived from HIF233A (chromosome 6) including the two parents as control, on a Copenhagen table as described previously (Khan et al 2012). HIF233A consists of 20 NILs and hundred randomly picked seeds from each harvest per NIL were imbibed in germination trays (21 × 15 cm; DBP Plastics NV, Antwerpen, Belgium, http://www.dbp.be) containing 50 mL of water and then stratified at 4°C for 3 days to break residual dormancy and to ensure uniform and rapid germination. After stratification the germination trays were transferred to a germination incubator set at 25°C and the seeds were allowed to germinate. Seed germination was recorded at 8h intervals and germinating seeds were transferred to the Copenhagen table. The first 40 germinated seeds per line were transferred to the Copenhagen table in four replicates of 10 seedlings per replicate in a randomized set up and the remaining seeds were discarded. The time to 50% germination of each replicate was recorded and the seedlings were allowed to grow on the Copenhagen table supplied with water without nutrients under long day conditions (16h light and 8h dark) at 25 °C. According to the schedule each plot was harvested upon completion of a 14d growth period from the time of completing 50% germination and the seedling phenotypes such as hypocotyl length and fresh shoot and root weight were recorded.

Phenotyping of seed traits

All the HIFs for chromosome 9 and their progenies were evaluated for seed quality traits (seed weight, dry seed size, dry seed length, imbibed seed area and imbibed seed length). For measuring the seed weight, on average 120 seeds were randomly taken from the seed lot of each individual and the weight of these seeds was measured on a sensitive balance and the average single seed weight (mg/seed) was calculated. The same weighed samples of the seeds were then transferred to germination trays lined with a white filter paper for taking close-up photographs using a Nikon D80 camera with a 60 mm objective fixed to a repro stand and connected to a computer, using Nikon camera control pro software version 2.0 (Joosen et al., 2010). After taking the photographs, the dry seeds were imbibed for 18h and the same procedure of taking close-up photographs was repeated for the imbibed seeds. The dry and imbibed seed size and length were determined by analysing the photographs 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 and seed length (SL) as the longest distance between any two points along the selection boundary (feret's diameter).

Statistical analysis of the data

Statistical analysis, including average, t-test and standard error were conducted to be able to detect QTL associations according to single-marker analysis (Collard et al., 2005). Progenies in each HIF were grouped for each segregating marker according to their background (Money, Pimp and heterozygous). A 2-tailed student's t-test was performed to analyse differences in mean values between groups carrying the parental allele (Money and Pimp) at 5% probability.

Results

Identification of seed and seedling trait related QTLs in the RIL population

QTL analysis of the phenotypic values for seed and seedling was carried out on the basis of the established marker linkage map of the *S. lycopersicum (*cv. Moneymaker, 'MM') × *S. pimpinellifolium* (G1.155, 'Pimp') RIL population, which consists of 865 SNP markers. As a result of several QTL analysis 62 QTLs were identified with moderate to large phenotypic effects (Khan et al., 2012) which were later on increased to 115 QTLs with an increase of the number of RILs from 83 to 100 (Chapter 3) for 20 seeds, seedlings and root system architecture(RSA) quality traits under both normal nutrients and nutrientless conditions. Highly significant levels of overlapping QTLs between phenotypic traits were also revealed by permutation tests conducted on all 1-LOD QTL intervals. Five QTL clusters positioned on

chromosomes 1, 4, 6, 9 and 11, affecting several seed dimensions, seedling biomass and RSA traits with overlapping proportions ranging from 62.5 to 100% of 1-LOD intervals were identified (Khan et al., 2012).

Development of near-isogenic lines

In the same RIL population HIFs were identified and constructed to evaluate 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). To identify families that were heterogeneous for the SNP marker most tightly associated with each QTL (Figure 6.1) we were able to develop HIFs. In order to identify each homozygous genotype at the region of interest and to study it in detail, the RILs that were still heterozygous at the position of the QTL were chosen, and then selfed and genotyped (F_8 plants). Several loci were investigated in order to find candidate RILs for the construction of the HIFs to confirm the QTL hotspots mapped on different chromosomes. We could not find HIFs to validate all the detected QTL clusters, but after screening the F_8 RILs we were able to find candidate RILs for the development of the HIFs for QTLs on chromosome 6 originating from 95233A (HIF233A) and 4 HIFs on chromosome 9 originating from 95233B (HIF233B), 95239 (HIF239), 95241 (HIF241) and 95259 (HIF259). These HIFs were screened with CAPS marker 43582592 on chromosome 6 whereas CAPS markers STW1070/1071, STW1000/1001, STW1054/1055 and STW1058/1059 were used to screen HIFs 233B, 239, 241 and 259 on chromosome 9. These HIFs have shown to segregate for the aforementioned markers on chromosomes 6 and 9. An example of selection and respective genotyping is presented for chromosome 6 in Figure 6.1.

Confirmation of seedling trait QTLs in near-isogenic lines

Confirmation of the QTLs on chromosome 6 for seedling traits

A number of significant occurrences of overlapping QTLs among seed and seedling traits were identified after permutation tests among all QTL positions in previous QTL mapping studies for seed and seedling quality phenotypes (Khan et al., 2012). One of the overlapping QTL hot spots was observed at the bottom of chromosome 6 where the confidence intervals significantly overlapped for a number of seed and seedling trait QTLs. Among others, the QTL for hypocotyl length was one of the most significant QTLs under both normal and nutrientless conditions at this locus. HIF233A showed a difference between MM and Pimp alleles in the region surrounding marker 42712640. The twenty NILs generated from HIF233A (chromosome 6), including the two parents, were tested for seedling traits, such as hypocotyl length, fresh shoot weight, fresh root weight, dry shoot

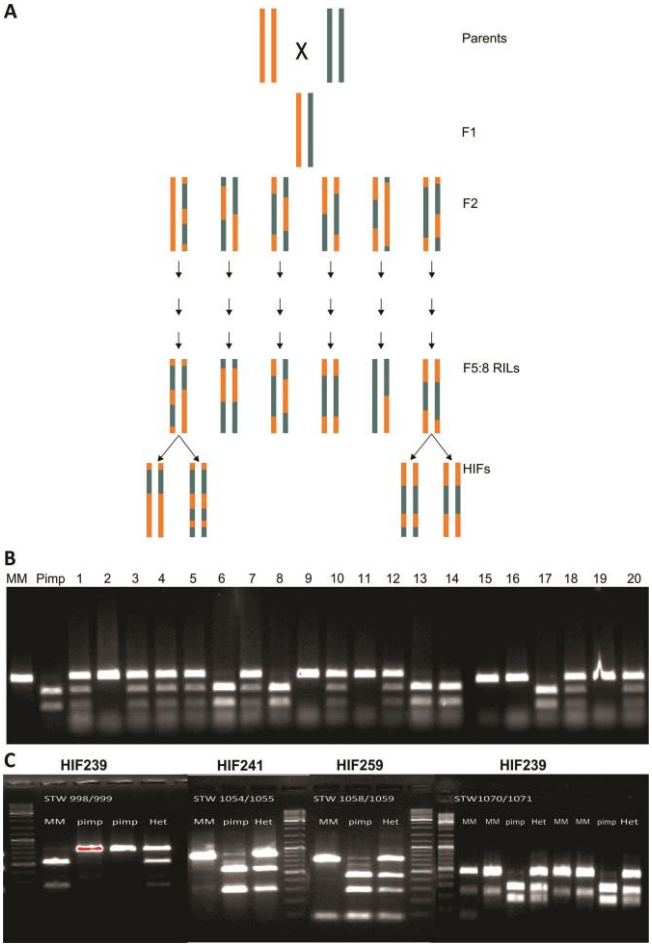


Figure 6.1. (A) Generating HIFs from a non-fixed recombinant inbred line by the methodology previously described by Tuinstra et al. (1997). The graphical genotype of individual plants is shown for a representative single pair of chromosomes. RILs can be produced by continued selfing of inbred lines that are not entirely homozygous until the F_7 generation (single-seed descent). Testing of such RILs with molecular markers around the QTL of interest enables the selection of heterozygous lines in this region. The phenotypic examination of progeny obtained by one additional further selfing in combination with further genotyping allows the selection of HIFs that are in a mixed but homozygous heterogeneous genetic background. Thus, pairs of NILs contrasting at marker loci connected with quantitative trait loci can be selected to obtain heterogeneous inbred 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 HIF233A are segregating for the marker; parents MM and Pimp are shown on the left. The HIFs having MM background are showing one band and the HIFs having Pimp alleles are showing two bands while the HIFs having heterozygous background are showing three bands. (**C**) The four markers used for HIFs on chromosome 9 are shown. Marker STW998/999 (HIF233B) shows two bands for MM background, one band for Pimp and three bands for Het background. Marker STW1054/1055 (HIF241) shows one band for MM, two bands for Pimp and three bands for Het background. Marker STW1058/1059 (HIF259) gives two bands for MM, three bands for Pimp and four bands for Het background. While marker STW1070/1071 (HIF239) reveals one band for MM, two bands for Pimp and three bands for Het background.

weight and dry root weight under nutrientless conditions. HIF 233A NILs were grouped into MM, Pimp and heterozygous (Het) backgrounds according to the segregating marker (42712640) at the locus of interest. Significant differences ($p<0.01$) were observed for hypocotyl length of the two parents as well as the NILs ($P<0.05$) having Money and Pimp backgrounds respectively (Figure 6.2). The average hypocotyl length (290 mm) of the NILs with the MM background was significantly higher ($P<0.05$) than the hypocotyl length (260 mm) of the NILs with Pimp background at the interval of interest. Thus, statistical evidence is given for the QTL that the MM allele is enhancing hypocotyl length at this locus. However, we could not confirm the QTLs for seed weight, fresh and dry root weight in the same interval on chromosome 6 (data not shown).

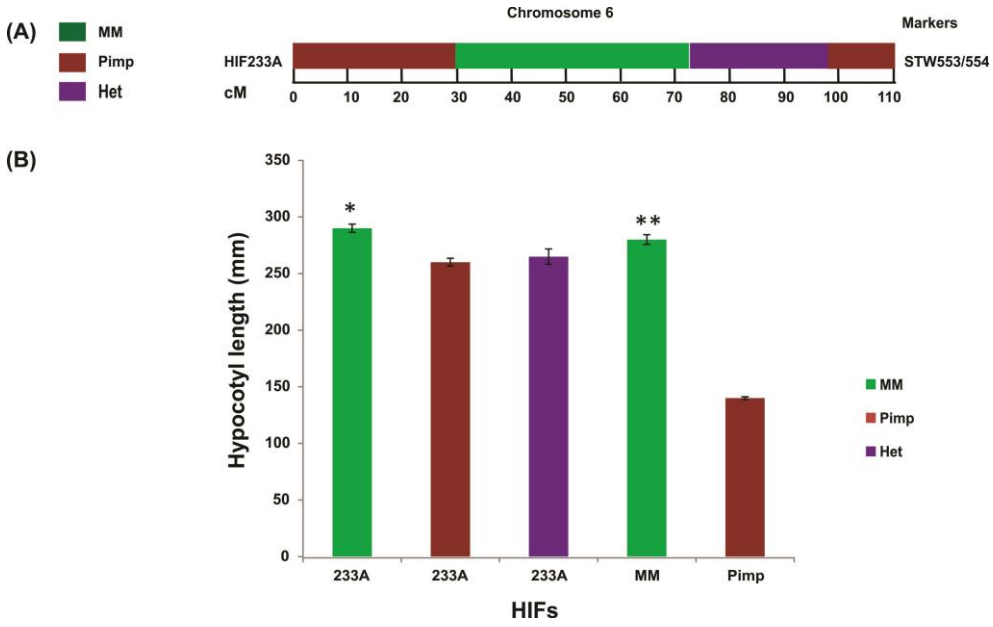


Figure 6.2. (**A**) Genetic map of the QTL on chromosome 6 for seed quality phenotypes. The different coloured boxes represent different genetic backgrounds. The heterozygous part for the HIFs (with blue colour) and the homozygous part for MM (green colour) and Pimp alleles (red colour) are shown. (**B**) Mean values for hypocotyl length in HIF233A homozygous for MM alleles (green), Pimp alleles (red) and heterozygous background (blue colour) and the two parental lines. *= significant difference ($p<0.05$), **= significant difference ($p<0.01$)

Confirmation of seed trait QTLs on chromosome 9 in near-isogenic lines

In our previous study (Chapter 2) we demonstrated that the genomic region of chromosome 9, ranging from 54.142 to 105.399 cM, contains a cluster of QTLs influencing seed and seedling traits. This region of chromosome 9 generally reveals one or two QTLs for most seed traits that are significantly overlapping with seedling trait related QTLs (Table 2.3 Chapter 2). Four HIFs (HIF 233, 239, 241and 259) showed a difference for MM and Pimp

alleles in the interval from 54.142 to 105.399 cM for markers STW1070/1071, STW1000/1001, STW1054/1055 and STW1058/1059. The generated HIFs were examined for different seed traits such as seed weight (SW), dry seed size (SS), dry seed length (SL), imbibed seed size (ImbSS) and imbibed seed length (ImbSL). The NILs were grouped into MM, Pimp and Het backgrounds according to the segregating markers in the region of interest.

Confirmation of seed weight QTLs

Figure 6.3 shows the results for seed weight of the four HIFs and the two parents. Highly significant differences were observed between the seed weight of the two parents. The MM parent had significantly heavier seeds (p value <0.001) compared to Pimp. Similarly, NILs from HIF241 and 259 showed significant (p value <0.01) differences for seed weight (Figure 6.3). In accordance to the QTL identified in QTL analysis where MM alleles added to the seed weight, the NILs of HIFs 241 and 259 having the MM background had significantly (p value <0.01) higher seed weight compared to the Pimp NILs. Thus, HIF241 and HIF259 provide statistical evidence of two separate seed weight QTLs in the intervals 78.009-92.585 and 94.46-112.29 cM as these two HIFs are not overlapping each other (Figure 6.3A). No significant differences in seed weight could be observed for the other HIFs.

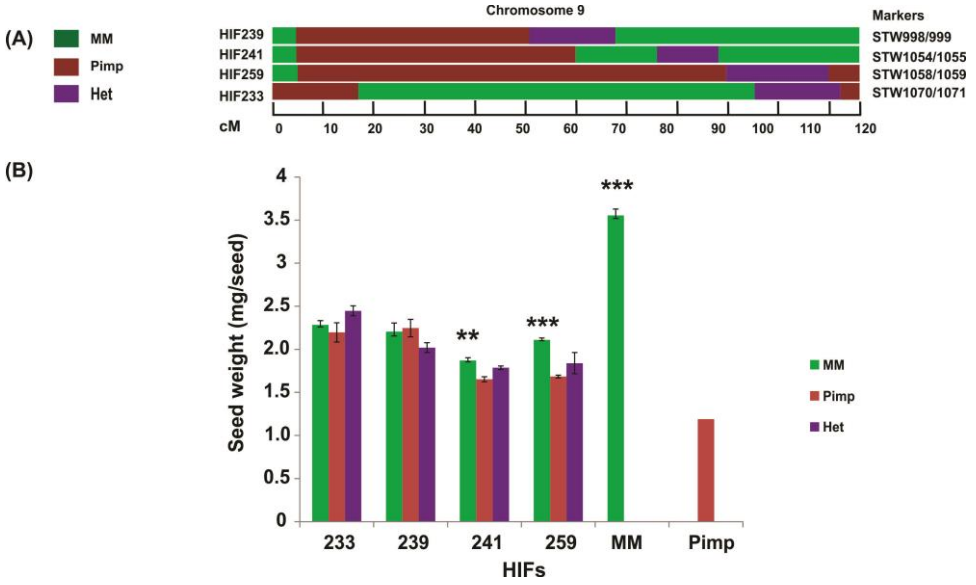


Figure 6.3. (**A**) Genetic map of the QTL on chromosome 9 for seed quality phenotypes. The different coloured boxes represent different genetic backgrounds. The heterozygous (Het) part for the HIFs (blue colour) and the homozygous part for MM (green colour) and Pimp alleles (red colour) on chromosome 9 are shown. (**B**) Mean values for SW in HIF233B, HIF239, HIF241and HIF259 homozygous for MM (green)and Pimp parent (red) and

heterozygous background (blue colour) including separate mean values for the two parental lines (MM and Pimp) are shown. **= significant difference (p<0.01), ***= significant difference (p<0.0001).

Confirmation of dry seed size (SS) QTLs

In addition to SW the SS of MM is significantly higher (p value < 0.001) as compared to the Pimp parent. Similarly, significant variation was also observed in the HIF lines homozygous for different parent alleles (Figure 6.4). In agreement with the SS QTLs identified in the QTL analysis where the MM parent alleles increased SS, the NILs of HIF259 and HIF233B having MM alleles in the interval of interest had significantly higher SS compared to the NILs homozygous for the Pimp alleles. Thus, these two HIFs provide evidence for the presence of QTLs for SS in the interval of 94.00-116.00 cM, as the two HIFs are overlapping in this interval. The variation in SS for the other two HIFs (HIF239 and 241) was not significant (Figure 6.4).

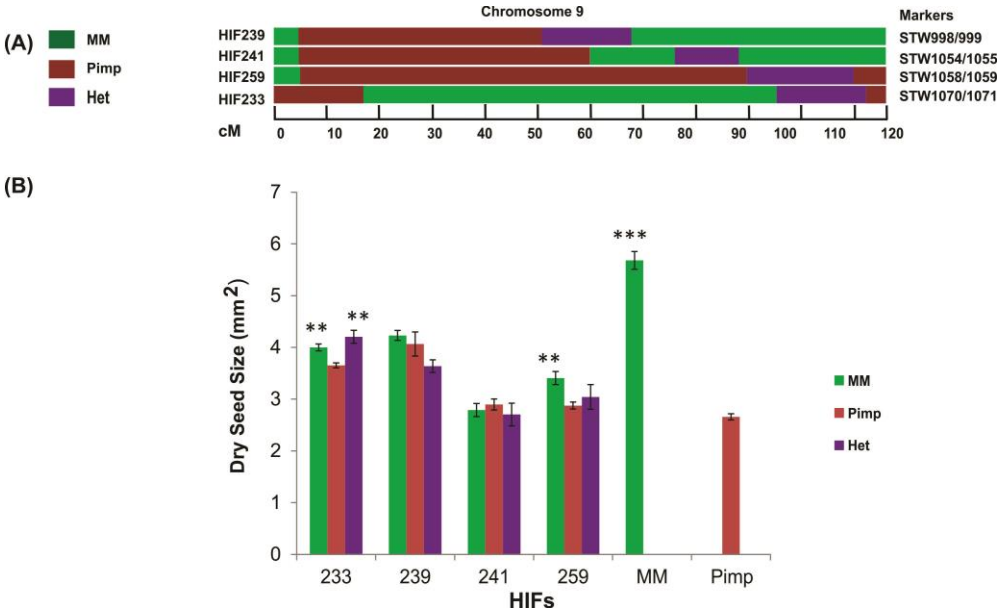


Figure 6.4. (A) Genetic map of the QTL on chromosome 9 for seed quality phenotypes. The different coloured boxes represent different genetic backgrounds. The heterozygous (Het) part for the HIFs (blue colour) and the homozygous part for MM (green colour) and Pimp allele (red colour) on chromosome 9 are shown. (B) Mean values for SS in HIF233B, HIF239, HIF241 and HIF259 homozygous for MM (green), Pimp (red) and Het background (blue) as well as the parental lines are shown. **= significant difference (p<0.01), ***= significant difference (p<0.0001).

Confirmation of dry seed length (SL) QTLs

The two parents showed significant differences in SL with the MM parent having significantly higher SL (p value <0.0001) as compared to the Pimp parent (Figure 6.5).

Comparing the dry seed length of the HIFs, statistically significant differences were observed among the NILs generated from HIF259 at the marker positions (Figure 6.5). NILs derived from HIF259 and having MM alleles revealed significantly higher (P <0.001) SL compared to NILs having Pimp alleles in the interval of this HIF. This is the only HIF that statistically confirmed the presence of a QTL for SL in this genomic region on chromosome 9, since no significant differences for SL were detected in the other 3 HIFs (HIF233, 239 and 241) (Figure 6.5).

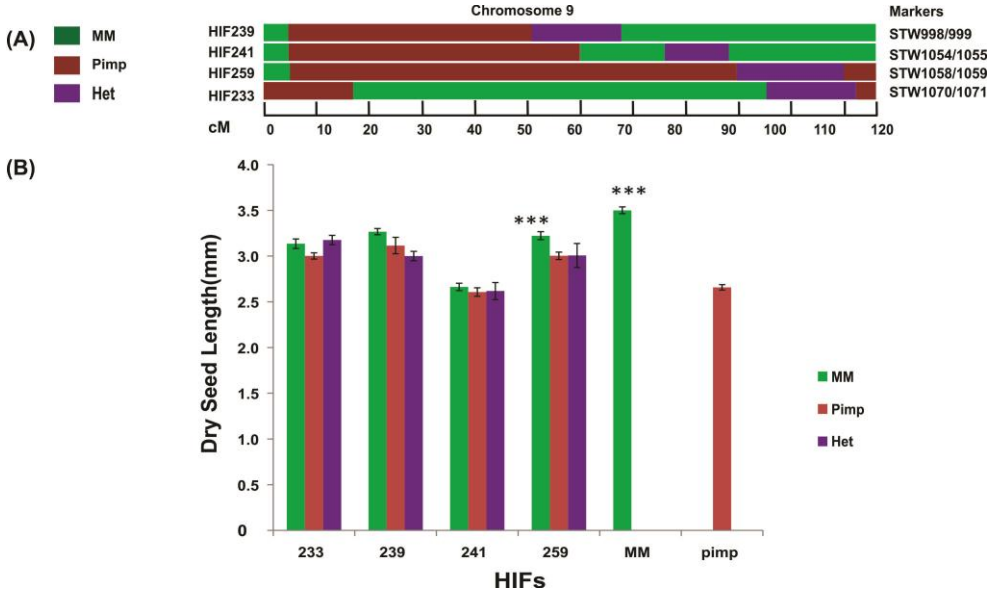


Figure 6.5. (**A**) Genetic map of the QTL on chromosome 9 for seed quality phenotypes. The different coloured boxes represent different genetic backgrounds. The heterozygous (Het) part for the HIFs (blue colour) and the homozygous part for MM (green colour) and Pimp allele (red colour) on chromosome 9 are shown. (**B**) Mean values for SL in HIF233B, HIF239, HIF241 and HIF259 homozygous for MM (green), Pimp (red) and Het background (blue) and the parental lines. **= significant difference (p<0.01), ***= significant difference (p<0.0001).

Confirmation of imbibed seed size (ImbSS) QTL on chromosome 9

In imbibed conditions seeds become bigger due to water uptake and it makes the visualization easier. Therefore, in addition to dry seed size, imbibed seed size was also measured for better understanding the difference in seed size. ImbSS was scored from images taken after 18h of imbibition. In addition to the difference in ImbSS of the two parents, highly significant differences in ImbSS were observed (p<0.001) in the NILs derived from HIFs 241 and 259 at the marker loci linked with these HIFs (Figure 6.6). Lines with the MM background showed larger ImbSS than the lines with the Pimp or Het background. Thus these two markers provide an evidence of presence of two separate QTLs for ImbSS in

the interval of the markers. This confirmation of the QTL is in agreement with the QTL for ImbSS where the additive effect was from the MM parent. The variation in ImbSS for the rest of the HIFs was not significant (Figure 6.6).

Confirmation of imbibed seed length (ImbSL) QTL on chromosome 9

During the image analysis, imbibed seed area, circularity and seed length are automatically scored. Therefore, the ImbSL was measured in one go with measuring the ImbSS from the same images of 18 hours imbibed seed. The two parents revealed highly significant (p value < 0 .0001) differences with the MM parent having a longer ImbSL as compared to Pimp. In accordance with the ImbSS the NILs generated from two HIFs (HIF241 and 259) displayed significant differences for ImbSL for the different genetic backgrounds. NILs with the MM alleles had significantly (p values ≤ 0.01) longer ImbSL than the NILs with Pimp alleles in the interval of interest (Figure 6.7). Thus these two HIFs confirm two ImbSL QTLs in the genomic intervals of 78.00-92.58 cM and 94.46-112.29 cM on chromosome 9. The other HIFs were non-significant as in case of ImbSS.

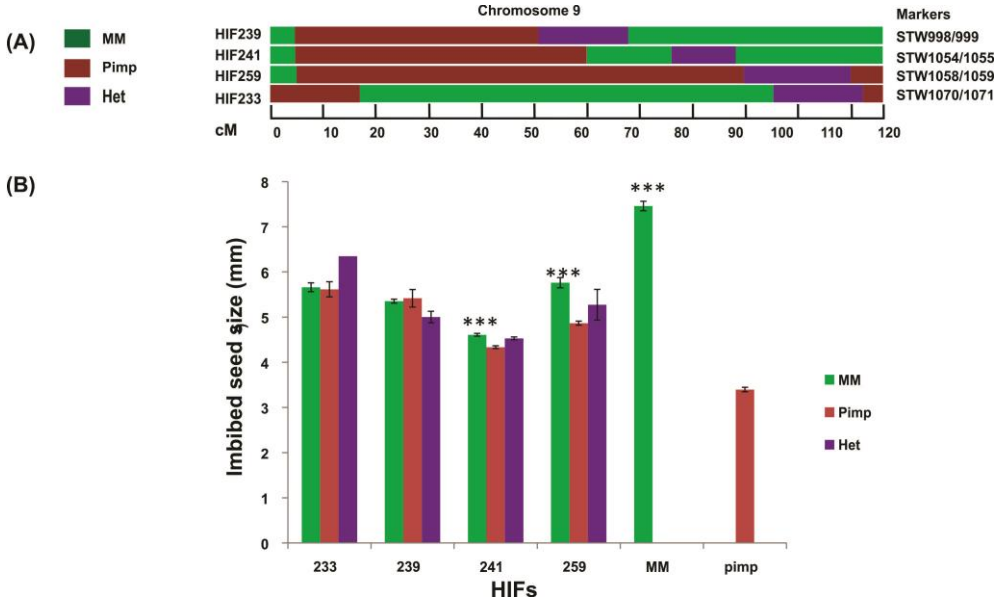


Figure 6.6. (**A**) Genetic map of the QTL on chromosome 9 for seed quality phenotypes. The different coloured boxes represent different genetic backgrounds. The heterozygous (Het) part for the HIFs (blue colour) and the homozygous part for MM (green colour) and Pimp allele (red colour) on chromosome 9 are shown. (**B**) Mean values for ImbSS in HIF233B, HIF239, HIF241 and HIF259 homozygous for MM (green), Pimp (red) and Het background (blue) and the parental lines. ***= significant difference (p<0.001).

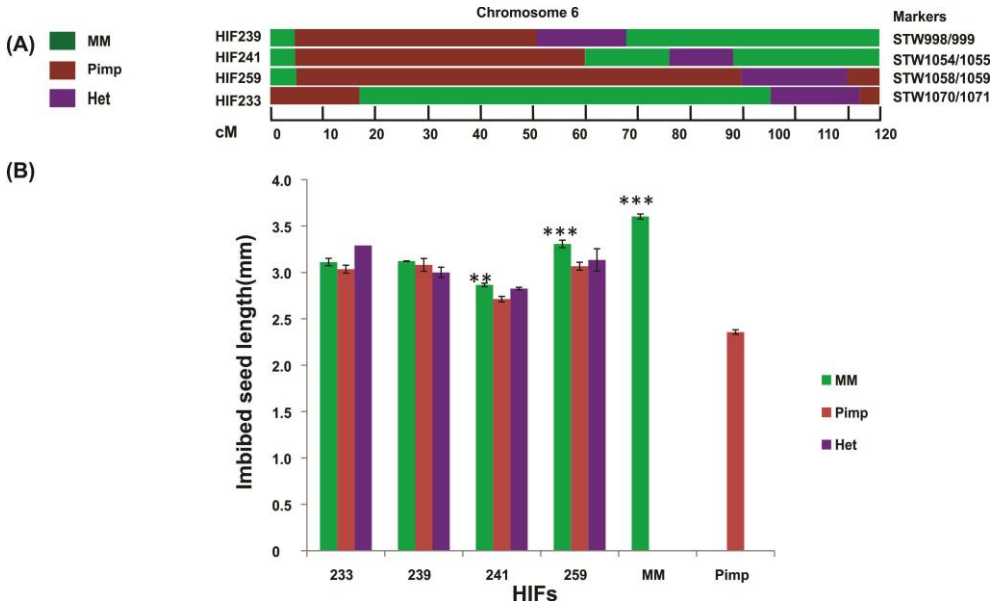


Figure 6.7. (**A**) Genetic map of the QTL on chromosome 9 for seed quality phenotypes. The different coloured boxes represent different genetic backgrounds. The heterozygous (Het) part for the HIFs (blue colour) and the homozygous part for MM (green colour) and Pimp alleles (red colour) on chromosome 9 are shown. (**B**) Mean values for ImbSL in HIF233B, HIF239, HIF241 and HIF259 homozygous for MM (green), Pimp (red) and Het background(blue) and the parental lines. **= significant difference ($p<0.01$), ***= significant difference ($p<0.001$).

Seed germination (%)

In addition to seed dimension traits we also tested the HIFs on chromosome 9 for seed germination capacity. The number of seeds germinated per line was scored 5d after imbibition to check if there was a variation in germination percentage in the HIFs. The Pimp parent revealed a significantly higher germination percentage compared to the MM parent. Although we observed significant differences among different HIFs for seed germination (%), these were not associated with seed traits or allelic variation within the HIFs as there was no significant difference between the different NILs derived from the same HIF (Figure 6.8).

Discussion

Depending on the availability of well-genotyped mapping populations, such as RIL populations, QTL analysis is a powerful tool to elucidate complex traits of the genetic differences that are present within a species (Koornneef et al., 2004). As RIL populations provide an 'immortal' genetic resource, many replicates of identical lines can easily be

studied in many different environments. However, identifying the molecular basis of QTLs remains a challenge, as QTL identification in this kind of analysis results in large genetic intervals, and thus requires molecular characterization of the allelic variation for cloning the corresponding genes. In order to reduce these QTL intervals and to fully understand the natural genetic variation for the identification of causal genes, one needs to pursue the molecular characterization of the identified QTL through integration of genome-wide information, which is becoming increasingly available.

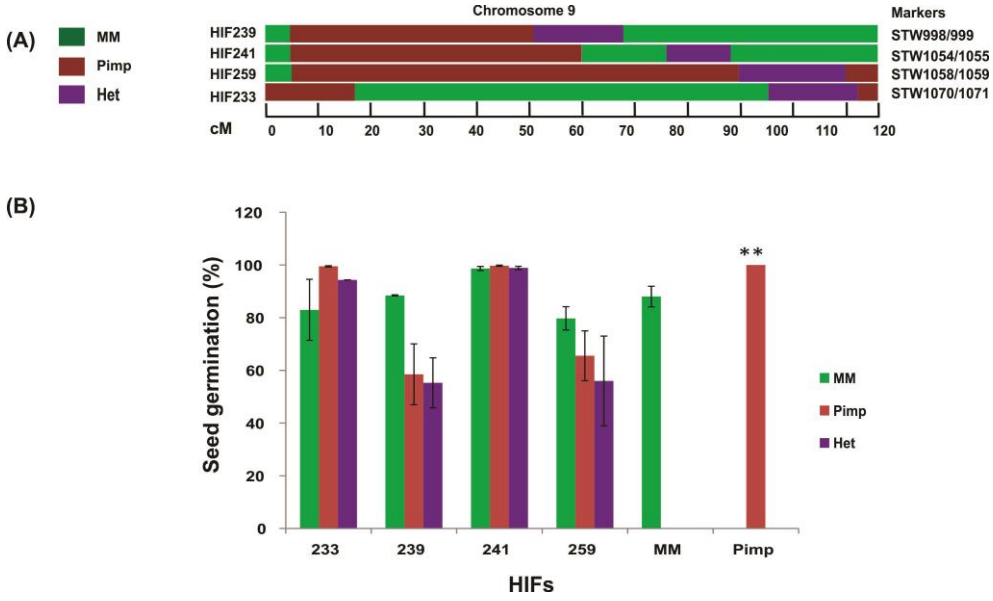


Figure 6.8 (**A**) Genetic map of the QTL on chromosome 9 for seed quality phenotypes. The different coloured boxes represent different genetic backgrounds. The heterozygous (Het) part for the HIFs (blue colour) and the homozygous part for MM (green colour) and Pimp allele (red colour) on chromosome 9 are shown. (**B**) Mean values for seed germination (%) of HIF233B, HIF239, HIF241 and HIF259 homozygous for MM (green),Pimp (red) and Het background (blue) and the parental lines. **= significant difference ($p<0.01$).

This molecular characterization of QTLs can be performed by the molecular isolation of the genes underlying the individual QTLs (the so called quantitative trait genes, QTGs) and the identification of DNA polymorphisms regulating the function of the gene and causing the phenotypic variation (referred to as quantitative trait nucleotides; QTNs). Further, the molecular characterization of individual QTLs needs their isolation from other segregating loci by constructing near isogenic lines (NILs) that differ only at a small region at the QTL of interest or at several closely linked loci affecting the trait of interest (Glazier et al., 2002). These NILs can be generated by recurrent backcrossing of RILs or, alternatively, NILs can be derived from a heterogeneous genetic background using HIFs. These HIFs are produced from plants heterozygous for a specific genomic region of interest in an otherwise

homozygous background (Tuinstra et al., 1997; Loudet et al., 2005). Compared to the construction of traditional NILs which requires several backcrosses, the HIF approach is less time consuming. In the present study NILs were generated 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 at the region of the QTL of interest.

In the current study, QTLs influencing seedling traits such as hypocotyl length fresh and dry root weight under nutrientless conditions at the bottom of chromosome 6 and seed traits at the second half of chromosome 9 were tested. The QTL for hypocotyl length on chromosome 6 was confirmed in HIF233A. NILs having MM alleles significantly increased hypocotyl length. This is in agreement to the additive effect of the original QTL identified for hypocotyl length at this position of chromosome 6 (Khan et al., 2012) where the MM alleles increase hypocotyl length. However, we could not confirm the QTLs for SW, ImbSL and other seedling traits at this position of chromosome 6.

In a previous study (Khan et al., 2012), QTLs for different seed phenotypes were often found in clusters across the 12 tomato chromosomes, in particular in the region of 50.0 to 115.0 cM on chromosome 9 where several seed and seedling QTLs were identified in a large cluster. These QTLs co-locating for dry and imbibed seed traits on chromosome 9 were characterized by testing four HIFs segregating for the region on chromosome 9, bearing several seed and seedling trait related QTLs. We confirmed that the variation observed in the previous phenotyping experiments in the RILs is relevant and repeatable. These four HIFs cover most of the region of interest (54 to 116 cM) on chromosome 9 in which the major QTLs for seed and seedling traits have been revealed in our previous study. The analysis of these HIFs confirmed the co-locating QTLs for five seed traits (SW, SS, SL, ImbSS and ImbSL). Seeds of HIF241 and 259 carrying MM alleles showed significantly higher SW compared to lines carrying Pimp alleles. These two HIFs are not overlapping with each other as HIF241 covers the genomic region ranging from 78.00 to 92.58 cM while HIF259 ranges from 94.46 to 112.29 cM. Thus, confirming two SW QTLs in this region of chromosome 9 is in agreement to our previous finding (Khan et al., 2012). The other two HIFs (HIF233B and HIF239) did not show any significant difference in SW for different parental alleles. HIF239 ranges from 45.93 to 70.85 cM while HIF233B covers the last part of the interval (98.74 to 115.20 cM). Thus the absence of variation in different allelic backgrounds for these two HIFs reduces the region of interest for the two SW QTLs to a smaller region ranging from 78.0 to 98.69 cM.

The seeds of HIF233B and the same HIF259 carrying MM alleles revealed larger SS as compared to the seeds carrying Pimp alleles. As these two HIFs are overlapping at the bottom of chromosome 9 in the interval of the SS QTL after 98 cM (Table 6.1 and Figure 6.4A), it might be possible that they indicate the same QTL for SS. On the other hand, seeds of HIF241 did not reveal significant differences for SS for the different genetic alleles, thus confining the interval of the SS QTL to a shorter region of 18 cM from 94 to 112 cM. Among

all the four HIFs, HIF259 is the most dominant, since the seeds carrying the MM alleles showing significantly higher phenotypic values for all the five tested seed traits (SW, SS, SL, ImbSS and ImbSL), followed by HIF241 which is significant for ImbSS and ImbSL in addition to SW. In addition, the seeds of HIF239 which cover the genomic region from 54 to 70 cM showed no statistical difference for the tested seed traits among the different genetic backgrounds and this eliminates the possibility of existence of any QTL for the tested seed traits in this interval.

In addition to seed dimension traits, these four HIFs on chromosome 9 were also validated for seed germination. Although significant differences were observed among different HIFs for seed germination (%), the difference between the NILs derived from the same HIF, carrying the different parental alleles, was not significant. This confirms the absence of QTLs for seed germination in the same interval as the QTLs for seed dimension traits. This supports our previous conclusion that although seed weight can be beneficial for seedling vigor, there is no association between seed size and seed germination performance (Khan et al., 2012). However, in case of HIF233A, in addition to the confirmation of the hypocotyl length QTL, several seed germination performance QTLs were confirmed in a previous study (Kazmi et al., 2013).

Confirmation of the QTLs controlling hypocotyl length on chromosome 6 and seed traits at different developmental stages (dry and 18h imbibed seeds) on chromosome 9 strengthens our previous findings of the presence of genetic regulation of seed germination (chromosome 6) and seed and seedlings traits (chromosome 9). Seed quality loci at this region (HIF233) of chromosome 6, in addition to hypocotyl length, also show relationships among the ability of seeds to germinate rapidly under different environments (Kazmi et al 2013). While the region of chromosome 9 in which we have developed HIFs (HIF233B, 239, 241 and 259) reveals greater genetic regulation of seed and seedling traits (Khan et al 2013). Interestingly, inside the QTL cluster on chromosome 9, the allelic effects are from the same parent which reinforces the possibility that one gene accounts for all the co-localizing QTLs. Hence, it could reasonably be hypothesized that the QTLs validated at the bottom of chromosomes 6 and 9 for seed and seedling traits are candidates that can be used in marker-assisted selection, or gene cloning by fine-mapping. Several other HIFs on other chromosomes are in preparation, especially on chromosomes 1 and 4, which can help in confirming and characterization of seed and seedling QTLs on these chromosomes identified in our previous study (Khan et al., 2012). Such confirmation of QTLs through HIFs has been documented for seed germination traits (Joosen et al., 2012) , for plant fructose and starch content (Calenge et al., 2006), early flowering time (Jiménez-Gómez et al., 2010) and root growth and architecture (Loudet et al., 2005) in Arabidopsis, as well as for disease resistance in maize (Chung et al., 2011) and flowering time in soybean (Su et al., 2010).

A good quality of HIFs is their genomic composition which, although homozygous, is a mixture of the two distinct parental lines. This provides the advantage that often more

than one HIF can be selected which offers the option to validate the same locus in distinct genetic backgrounds. This allows the elucidation of QTLs for epistatic interactions with other genomic regions, which otherwise can only be achieved by crossing pure introgression lines (Loudet et al., 2005).

Our present study has revealed the differing genetic control of seed traits and hypocotyl length in the MM and Pimp genetic backgrounds. However, the major QTLs for seedling traits on chromosome 1, 4, 9 and 11 remain to be confirmed in the HIFs developed on these chromosomes, but not yet investigated. Continuous variation for seed and seedling development (the RIL population) as well as discrete variation in different backgrounds (the HIF lines) is available to test these hypotheses.

The integration of these confirmations with the previous identification of QTLs provides support for continuing further fine mapping and cloning of gene(s) underlying the complex trait variation, which usually requires the construction of NILs in order to determine the exact genetic relationship among these traits. Hence, further molecular characterization of these candidates, including quantitative expression analysis in HIF lines can be combined with a candidate gene approach by looking at all the genes in the seggregating interval to assess the mechanisms involved in genetic variation for the complex traits. The investigation of QTLs for seed and seedling traits in tomato revealed that HIF analysis could identify NILs that allow the confirmation of linkage between markers and QTLs, for fine mapping of QTLs, and for elucidating the phenotype associated with a particular QTL. One of the drawbacks in our study was the limited number of NILs derived from the HIFs, especially on chromosome 6, due to which we could not confirm the small effect QTLs for seed and seedling traits such as fresh and dry root weight at the same position where we confirmed the QTL for hypocotyl length. This problem can be resolved by increasing the number of HIFs, by detecting recombination in the interval, by screening the progenies of heterozygous plants of each HIF, and to use these, once fixed as MM and Pimp alleles, as new HIFs with smaller candidate regions. Furthermore, the detailed analyses of genetic architecture requires consideration of multiple populations that represent a larger sample of the standing genetic variation in the species and thus provide a framework for comparative analyses (Collard and Mackill, 2008).

In the present study the analysis of HIFs reveals an efficient approach to develop NILs for dissecting the genetic basis underlying seed and seedling phenotypes. This HIF analysis provides additional evidence for the presence of genetic regions previously identified to control seed quality. Studying different seed and seed phenotypes is particularly important for improving crop growth and final yield of the crop. HIFs can help in generating NILs in a range of recombinant genetic backgrounds and are useful in confining the genetic regions in which the phenotype of a QTL is clearly expressed. This may facilitate fine mapping and cloning of these QTLs and could thus lead to the identification of genes potentially involved in the control of different linked physiological processes.

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

Alpert KB, Tanksley SD (1996) High-resolution mapping and isolation of a yeast artificial chromosome contig containing fw2. 2: a major fruit weight quantitative trait locus in tomato. Proceedings of the National Academy of Sciences of the United States of America **93:** 15503-15507

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

Barton N, Turelli M (1989) Evolutionary quantitative genetics: how little do we know? Annual Review of Genetics **23:** 337-370

Borevitz JO, Chory J (2004) Genomics tools for QTL analysis and gene discovery. Current Opinion in Plant Biology **7:** 132-136

Calenge F, Saliba-Colombani V, Mahieu S, Loudet O, Daniel-Vedele F, Krapp A (2006) Natural variation for carbohydrate content in Arabidopsis. Interaction with complex traits dissected by quantitative genetics. Plant Physiology **141:** 1630-1643

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

Chung C-L, Poland J, Kump K, Benson J, Longfellow J, Walsh E, Balint-Kurti P, Nelson R (2011) Targeted discovery of quantitative trait loci for resistance to northern leaf blight and other diseases of maize. Theoretical and Applied Genetics **123:** 307-326

Clouse SD, Sasse JM (1998) Brassinosteroids: essential regulators of plant growth and development. Annual Review of Plant Biology **49:** 427-451

Collard BC, Mackill DJ (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philosophical Transactions of the Royal Society B: Biological Sciences **363:** 557-572

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. Theoretical and Applied Genetics **112:** 139-148

Eshed Y, Zamir D (1995) An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. Genetics **141:** 1147

Eshed Y, Zamir D (1996) Less-than-additive epistatic interactions of quantitative trait loci in tomato. Genetics **143:** 1807-1817

Fridman E, Pleban T, Zamir D (2000) A recombination hotspot delimits a wild-species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene. Proceedings of the National Academy of Sciences of the United States of America **97:** 4718-4723

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

Holland JB (2007) Genetic architecture of complex traits in plants. Current Opinion in Plant Biology **10:** 156-161

Jansen R, 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

Jiménez-Gómez JM, Wallace AD, Maloof JN (2010) Network analysis identifies ELF3 as a QTL for the shade avoidance response in Arabidopsis. PLoS Genetics **6:** e1001100

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. Plant Journal **62:** 148-159

Joosen RVL, Arends D, Willems LAJ, Ligterink W, Jansen RC, Hilhorst HW (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. Theoretical and Applied Genetics **87:** 233-237

Kazmi R, Willems L, Ligterink W, Hilhorst H (2013) Dissection of the complex phenotypes of seed quality on tomato chromosome 6 and 8: HIFS (Heterogenous inbred families) confirmation. PhD Thesis. Genes for seed quality**:** 179

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

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 JJ (2012) Backcross Populations and Near Isogenic Lines. *In* Quantitative Trait Loci (QTL). Springer, pp 3-16

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

Kroymann J, Textor S, Tokuhisa JG, Falk KL, Bartram S, Gershenzon J, Mitchell-Olds T (2001) A gene controlling variation in Arabidopsis glucosinolate composition is part of the methionine chain elongation pathway. Plant Physiology **127:** 1077-1088

Ligterink W, Joosen RV, Hilhorst HW (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. 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. Theoretical and Applied Genetics **110:** 742-753

Mackay T (1995) The genetic basis of quantitative variation: numbers of sensory bristles of *Drosophila melanogaster* as a model system. Trends in genetics **11:** 464

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

Rikke BA, Johnson TE (1998) Towards the cloning of genes underlying murine QTLs. Mammalian Genome **9:** 963-968

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

Ron M, Weller J (2007) From QTL to QTN identification in livestock–winning by points rather than knock-out: a review. Animal Genetics **38:** 429-439

Schlichting CD, Pigliucci M (1995) Gene regulation, quantitative genetics and the evolution of reaction norms. Evolutionary Ecology **9:** 154-168

Shrimpton A, Robertson A (1988) The isolation of polygenic factors controlling bristle score in *Drosophila melanogaster*. II. Distribution of third chromosome bristle effects within chromosome sections. Genetics **118:** 445-459

Slate J (2005) Quantitative trait locus mapping in natural populations: progress, caveats and future directions. Molecular Ecology **14:** 363-379

Su C, Lu W, Zhao T, Gai J (2010) Verification and fine-mapping of QTLs conferring days to flowering in soybean using residual heterozygous lines. Chinese Science Bulletin **55:** 499-508

Sweeney M, McCouch S (2007) The complex history of the domestication of rice. Annals of Botany **100:** 951-957

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. 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. 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. 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.

Wu R, Lin M (2006) Functional mapping-how to map and study the genetic architecture of dynamic complex traits. Nature Reviews Genetics **7:** 229-237

Chapter 7

General Discussion: Evaluation of Seed Quality Through Systematic Genetic and Physiological Approaches

Seed Quality

Seeds are the major means of regeneration of plants as well as sources of food, feed, fibre, biofuels and bioactives and their value lies mainly in the storage reserves of protein, starch and oil which are synthesized during seed development and maturation. Seed must constantly be propagated, harvested and after harvesting processed and stored correctly and carefully in order to maintain and maximize its vigour and viability for subsequent crop productivity. Seed vigour denotes the ability of a seed lot to uniformly germinate to its maximum percentage under a wide range of environmental conditions, and to establish healthy seedlings. Seed quality can be assessed in several ways, including testing for genetic and physical purity, mechanical damage and physiological attributes, such as viability, germination, dormancy, vigour and uniformity (Dickson, 1980; Hilhorst and Toorop, 1997; Hilhorst, 2007, 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, as well as seed treatments (harvesting and processing) contribute to the 'accumulated damage' that influences 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 at the various stages of the seed life (Koornneef et al., 2002) and, therefore, can be challenged over the entire seed production chain.

The seed is the means by which the new individual is dispersed. Thus the seed occupies a critical position in the life history of higher plants. The success with which the new individual is established and the vigour of the young seedling are largely determined by the physiological and biochemical features of the seed. Of key importance to this success are both the responses of a seed to its environment and the food reserves it contains; these are available to sustain the young plant in the early stages of growth before it becomes an independent, autotrophic organism, able to use light energy.

Natural Variation for the Complex Trait of Tomato Seed Quality

Seed quality is an important agronomic and quantitatively inherited trait, yet very little is known, at the molecular/genetic level, about the final establishment of seed quality. In the present seed production practice the emphasis is on harvest and post-harvest treatments since it is difficult to influence the production environment, even under greenhouse conditions, where influences from the outside can also not be avoided. The genetic components of seed quality have hardly been used in breeding programs. Some quantitative trait loci (QTLs) related to germination, storability and stress resistance have been found in Arabidopsis and tomato (Foolad et al., 2003; Clerkx et al., 2004; Joosen et al., 2012) but a systematic study of the genetics of seed quality is lacking. With the current technological progress it has become possible to combine quantitative genetics with genomics approaches and this thesis aims at dissecting the genetic of seed quality by integrating genetics with detailed phenotyping at the physiological, metabolomic and genomic level to understand underlying genetic mechanisms regulating various attributes of seed and seedling quality.

Exploiting the natural variation present in a *S. lycopersicum* x *S. pimpinellifolium* Recombinant Inbred Line (RIL) population with respect to seed and seedling phenotypes provides an opportunity to dissect and understand the physiological and genetic mechanisms governing seed and seedling quality (Alonso-Blanco and Koornneef, 2000). Such a systematic approach, studying genetic populations differing in seed-and seedling-quality parameters, may provide valuable insight into the involvement of genes, and the processes they control, in the acquisition of seed quality (Ligterink et al., 2012). To date, only a few QTL positions have been cloned and characterized in detail, but if genes or gene sets associated with seed and seedling quality parameters become available, they may be used as diagnostic tools to evaluate seed quality, in marker-assisted breeding, or in genetic modification to enhance seed quality. Thus, to link bio-molecular and phenotypic variables in order to identify novel alleles for tomato seed quality at various levels of organization, are highly desirable. Therefore, to reduce the gap between knowledge and practice and create meaningful and significant biomarker- phenotype complexes, we report on a comprehensive study that applies an integrated approach to assess the attributes of seed quality and its association with seedling phenotypes.

Genetic Analysis of Seed Dimensions and their Link with Seedling Establishment

Seed size is a key determinant of evolutionary fitness in plants and is a trait that often undergoes tremendous changes during crop domestication. Seed size is most often

quantitatively inherited, and there is a very large environmental influence on seed size with many factors that interact to affect the trait (Fenner, 1991). Nevertheless, the seeds of domesticated plants are normally much larger than those of their wild counterparts and this suggests that this change in seed weight was likely the result of selection pressure for yield, improved germination and seedling vigour during the process of domestication (Doganlar et al., 2000). However, in many cases there is no clear link between seed size and yield as there are trade-offs between these and other components of yield such as seed number (Alonso-Blanco et al., 1999; Mahmood et al., 2005). However, seed size also increased during domestication in crops other than the one harvested for their edible seed. An example, is the domesticated tomato that produces seeds up to several fold larger than its wild ancestors (Doganlar et al., 2000). Why seed size increased during domestication in crops not consumed for their seeds is unknown. However, it is generally believed that seed size increased in these species due to indirect selection for greater seedling vigour and germination uniformity under field production (Harlan et al., 1973). In tomato, seed size is also positively correlated with fruit size (Goldman et al., 1995; Grandillo and Tanksley, 1996) which could be one of the reasons for selecting unintentionally for larger seeds. However, despite the evolutionary and agronomic significance of seed weight/and or size, relatively little is known about the genetic and molecular processes underlying natural variation in seed size and, thus far, none of the genes in this pathway have been isolated from any plant species, although Orsi and Tanksley (2009) have cloned *sw4.1* for seed weight but have not yet identified the causal gene. Most of our knowledge comes from quantitative trait loci mapping studies which have revealed a fairly large number of QTLs affecting seed size in a variety of plants including Arabidopsis, cowpea, mungbean, green pea, soybean and tomato (reviewed by Doganlar et al., 2000).

The seeds of tomato are composed of an embryo, an endosperm and the seed coat. Each of these three structures is genetically distinct and could potentially contribute to seed-weight variation. The seed weight of tomato is quantitatively inherited and determined mainly by additive gene action (Nieuwhof et al., 1989). Tomato is one of those species not domesticated for edible seeds, but in which extensive QTL mapping for seed weight has been conducted over the past 28 years. These studies, involving crosses between the cultivated tomato and its related wild forebearers, have identified approximately 20 QTLs related to seed weight variation (Doganlar et al., 2000). Different subsets of these QTLs were identified in different studies. However, in most of these QTL mapping studies seed dimensions, such as size, length and circularity as well as imbibed seed size, length and circularity were ignored but these are important attributes of seed size. In our current QTL mapping study, in addition to seed weight, we also investigated the dimensions of the seed in both dry and 18h imbibed seeds and identified 42 genomic regions (QTLs) regulating seed dimensions, which is an important contribution to our understanding of the genetic components regulating seed weight variation in tomato. A

very clear link between seed size and the resulting size of the seedling, which is dependent on seed reserves for pre-emergence growth, has been established in *Brassica oleracea* (Finch-Savage et al., 2010) and other crops (Bettey et al., 2000). Large seeds tend to produce larger seedlings at emergence and this size difference can be maintained to harvest (Benjamin, 1990; Fenner, 1991). However, no attention in any of the above mentioned QTL mapping studies has been given to this aspect of seed weight in tomato. In our current study we exploited the natural variation present in the RIL population of *S. lycopersicum x S. pimpinellifolium* for seed traits and documented a strong physiological correlation and pleiotropic co-location of QTLs between seed dimensions and seedling quality traits. These physiological correlations and pleiotropic co-location of QTLs between seed and seedling traits have established a strong association between seed weight/size and seedling vigour related traits.

Seedlings are living, perishable plants and their establishment is determined by interacting genetic, physiological and environmental components (Hodgkin and Hegarty, 1978; Wright and Westoby, 1999). Good seedling establishment is essential for crop production to be both sustainable and profitable and is therefore widely accepted as a critically important trait for farmers. 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 seedling's ability for shoot penetration through the impeding soil of the seed bed is an essential attribute of vigour (Khan et al., 2012). The ability to germinate faster, followed by seedling growth, represents key phenotypic markers for seed vigour 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 vigourous seed should possess the ability to establish healthy seedlings across various environments. Seedling shoot height and 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 for predicting seed vigour (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 have been ignored in previous studies (Chapter 2). Although seed size, length and weight are interdependent traits, this study also showed significant differences in the total number and locations 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 2, found strong relationships between the various 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 vigour) on the initial downward growth of the root system has been reported, 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 could be due to common genetic mechanisms controlling these traits (Khan et al., 2012). 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).

Genetic Analysis of Germination Phenotypes

In addition to seed size and seedling establishment the natural variation present in a *S. lycopersicum* x *S. pimpinellifolium* recombinant inbred line (RIL) population was also explored for seed performance. The RIL population obtained from these accessions proved to be a powerful resource for the detection of seed quality QTLs related to seed germination under both controlled and various stress conditions. The QTL mapping approach appears to be valuable not only in elucidating the genetics, but also the physiological background of seed quality phenotypes. Seed germination begins with water uptake by the seed (imbibition) and completes with the radical protruding the endosperm and seed coat (Bewley, 1997). In Chapter 3 we 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 reveals the overall germination potential of 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). 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 ($t_{50}^{-1}$and t_{10}^{-1}, respectively, and MGR = mean germination rate), as well as uniformity (U_{7525}^{-1}). Thus, it was important to include the various aspects of cumulative germination in order to quantify the different seed quality traits under 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 the various 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 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 (Chapter 3; Kazmi et al., 2012). 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 suggest that the same genes contribute to the germination response under these various stress conditions. Thus, selection for rapid germination at one stress condition would result in progeny with improved germination under additional 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). It is well known that seed lots of similar viability, determined in standard germination tests (Finch-Savage et al., 2010) can perform differently under more stressful conditions due to differences in their vigour. A widely accepted definition of 'vigour' is the sum total of all those properties that enables the germination and emergence of seedlings under a wide range of environments (Perry, 1984). Vigour is therefore an estimate of how successfully a seed lot will establish seedlings under the wide range of conditions experienced in practice and this is determined by interacting genetic and environmental components (Whittington, 1973; Hodgkin and Hegarty, 1978). These components influence both seed germination and subsequent post-germination reserve-dependent seedling growth leading to emergence from the soil. 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 involved in acquisition of seed quality.

Genetic Analysis of Seed Major Reserves and their Link with Seed and Seedling Biomass

Seed reserves are of major importance as they support early seedling growth when degraded upon germination and, therefore, participate in crop establishment (Baud et al., 2002). The storage compounds that mostly accumulate during the seed filling phase are storage proteins (e.g., albumins, globulins and prolamins), oil (often triacylglycerols) and carbohydrates (often starch; Bewley et al., 2012). Seed weight is an indication of the amounts of reserves that seeds may contain and large seeds may establish vigorous seedlings supported by the larger amounts of reserve food that heavy seeds contain

(Wright and Westoby, 1999). The relationship between seed size and seed reserves has been reported and discussed in several studies (Lowe and Ries, 1973; Ries and Everson, 1973; Evans and Bhatt, 1977; Saxena et al., 1987; Weschke et al., 2000; Cui et al., 2002; Panthee et al., 2005), but with the exception of a few, no systematic study using genetic populations (RILs or ILs) is available. In our study (Chapter 2; Khan, et al., 2012), we exploited the natural variation present in the *S. lycopersicum* x *S. pimpinellifolium* RIL population for seed dimensions and seedling growth and demonstrated a strong correlation between seed dimensions, seedling biomass and root architecture and most of the QTLs for seed dimensions were co-locating with QTLs for seedling biomass under both normal and nutrient deficient conditions.

Based on these findings it was postulated that seedling vigour from larger seeds was related to higher amounts of seed reserves in these seeds. To test this hypothesis, it was important to investigate the RIL population for seed food reserves to find its association with seed and seedling biomass. Higher amounts of protein in larger seeds and the effect of genotype and its interaction with the environment and a positive link with seed and seedling biomass have been proposed and discussed in several studies (Lowe and Ries, 1973; Ries and Everson, 1973; Evans and Bhatt, 1977; Saxena et al., 1987; Panthee et al., 2005; Thomas et al., 2009). This was further confirmed in the present study (Chapter 4) where we demonstrated that larger seeds have significantly higher amounts of protein (both relative and total protein content: RAP and TPS respectively)as compared to small seeds. This was obvious from the RAP and TPS of the two parents of the RIL population as the *S. lycopersicum* parent, which has the larger seed weight/size contained significantly higher (29%) RAP and nearly four-fold higher TPS as compared to the small-seeded *S. pimpinellifolium* parent. Further, there was a consistent increase in the level of RAP and TPS with an increase in seed weight of the RILs in the population. In addition, a moderate correlation between RAP and seed/seedling biomass and a strong correlation between TPS and seed/seedling and root system architecture (RSA) traits was established. These correlations were further supported by a complete co-location of QTLs between RAP, TPS and seed/seedling and RSA vigour related traits. The correlation was stronger between TPS and seed and seedling traits in general and both RAP and TPS with seedling vigour related traits grown under stressed nutritional conditions, in particular. Similar findings have been reported in other studies where the relative amount of protein was positively correlated with seed/seedling properties but the highest correlation was demonstrated between the total amount of protein per seed and seed/seedling biomass (Lowe and Ries, 1973; Ries and Everson, 1973; Weis, 1982; Panthee et al., 2005).

With respect to starch, there was no significant difference in the relative amount of starch (RAS) of the two parents but the RIL population showed significant differences in their RAS due to transgressive segregation. However, the increase or decrease in the RAS of the RILs did not show the same trend as seed weight or size. This suggests that the amount

of starch and level of seed weight are under control of different genetic mechanisms. This was also confirmed by the lack of any significant correlation between RAS and seed/seedling biomass. Nevertheless the total amount of starch (TSS) of *S. lycopersicum* was significantly higher than of *S. pimpinellifolium* and the same was true for other seed size groups in the RIL population. This reveals that even if the level of starch remains the same, yet, the total amount of starch per seed is related to seed size. This was also obvious from significant positive correlation between TSS and most seed/seedling vigour related traits.

Other studies, depending on type and species of the studied crops, have reported both positive and negative correlations between seed starch content and seed/seedling biomass. It has been shown in several studies that in tomato seed size/weight is positively correlated with fruit size/weight and negatively with fruit soluble solid contents (Goldman et al., 1995; Fulton et al., 1997; Fulton et al., 2000; Monforte and Tanksley, 2000; Frary et al., 2004; Schauer et al., 2005; Prudent et al., 2009). No reports are available on the starch content of tomato seeds. However, in other crops both positive and negative correlations between seed starch content and seed and seedling growth have been observed. For example, in rice the starch content of the grain was reported significantly positively correlated with seed/seedling biomass as well as with root architecture (Cui et al., 2002). On the other hand, (Hicks et al., 2002) documented a negative correlation between seed starch content and seed weight in sorghum. Thus the correlation between individual seed reserves (protein, starch, and oil) and seed/seedling quality depends on the composition (relative abundance) of the reserves in the seed as well as on the type of crop. For example, it has been reported and reviewed that dicots such as legumes generally store higher levels of protein (21-40%) and oil as compared to starch, whereas most monocot seeds accumulate higher levels of starch and low levels of both protein and oil (Bewley et al., 2012). Being a dicot, tomato seed also stores higher levels (22-33%) of protein and lipid (20-29%) and low levels of starch (Schauer et al., 2005; Sheoran et al., 2005). This concept is further corroborated by our finding that irrespective of genotype and seed size, significantly higher levels (12-19 times) of protein were observed, as compared to starch in tomato seeds. Thus in tomato the strong phenotypic correlation of seed reserves (RAP, TPS and TSS) with seed/seedling and RSA traits establish a close physiological link between seed reserves and seed and seedling phenotypes. A perfect pleiotropic co-location of QTLs for RAP and TPS, as well as several RAS and TSS QTLs, with seed and seedling biomass QTLs also suggests a genetic link among them. However, a perfect co-location of RAP and TPS related QTLs with the QTLs for seed mass and seedling vigour proposes a greater role for seed protein content in seedling growth than starch content.

After we confirmed a close and positive association between seed weight/size and seed reserves and seedling morphology, we further investigated the relationship between seed and its major tissues.It has been reported that seed size or weight can prominently be

affected by the genotype of three different seed tissues of tomato: the maternal testa, the triploid endosperm and the diploid embryo (Doganlar et al., 2000; Orsi and Tanksley, 2009). Our study revealed that 70-80% of the seed size is explained by embryo size whereas endosperm and testa are comparatively thin and contribute less to seed size. On the other hand, the combined contribution of testa and endosperm to seed weight was higher (65-70%) as compared to embryo (30-40%). These observations are in the same range as documented by Sheoran et al. (2005), who reported that in tomato seeds (*S. lycopersicum*) the embryo accounts for about 35% of the total seed weight. A strong correlation between seed size/weight and embryo size/weight as well as endosperm weight suggests a strong interdependency between each other which might be genetically interlinked. Similar correlations within embryo and endosperm masses and between seed mass and dimensions have been revealed by Zhang and Maun (1993) in *Calamovilfa longifolia*. Positive correlations between endosperm mass and grain weight/size have been reported in a number of cultivars of field grown barley (Cochrane and Duffus, 1983). It has been reported in three strains of subterranean clover that embryo weight accounting for 66% of the seed weight was strongly correlated with seed weight (Black, 1957). Positive correlations between embryo and seed weight have been reported and discussed in pea (Davies, 1975), soybean (Egli et al., 1981), maize (Reddy and Daynard, 1983) and wheat (Jenner et al., 1991). Thus, there is evidence to support our finding that there are tight links between embryo, endosperm and seed size. For example Orsi and Tanksley (2009), found that an ABC transporter controls seed size in tomato through gene expression in the developing embryo. A synonymous increase in seed, embryo and endosperm size during different seed developmental stages has been reported with no difference in the ratio of increase in embryo and endosperm of small and large seeded tomato lines. Despite this correlation, embryo and endosperm are genetically different tissues and different genes or differential expression of genes have been described in these two tissues of the seed (Penfield et al., 2006).

Nevertheless, despite their close relationship, the embryo generally correlated more strongly with seed and all seedling and RSA traits, as compared to the endosperm of which correlation as a whole was lower and even absent in some seedling traits (e.g., hypocotyl length). This observation makes the two tissues different from each other in their contribution to seed quality and seedling vigour.

In addition to their size variation, differential accumulation of reserves has been reported in embryo and endosperm. Dicots usually accumulate higher amounts of protein and oil, mainly in the embryo, rather than starch while monocots, such as cereals generally store higher levels of starch mostly in the endosperm (Bewley et al., 2012). Our current study further reinforced this hypothesis as our results demonstrated that regardless of genotype or seed size/weight the RAP and TPS of the embryo were significantly higher as compared to RAP and TPS of the endosperm. Our data further reveals that the RAP and TPS

of both embryo and endosperm of larger seeds are significantly higher than in the small seeds. This was also in agreement with whole seeds where larger seed contained significantly higher amounts of protein per seed, as compared to small seeds whereas the RAS and TSS of the endosperm revealed the same trend as the RAS and TSS of the whole seeds of the two parents. The RAS of the endosperm was not significantly different but the TSS of the larger seeds was significantly higher. Furthermore, our results show that the RAS of embryo of the small seeds was significantly higher related to the RAS of embryo of larger seed. Further differences in the two seed tissues were revealed in the correlation analysis where the RAP of the embryo was significantly correlated with all the seed dimensions, seedling and RSA traits whereas those of the endosperm did not correlate with most of the RSA traits. This is making the embryo more specialized for reserve accumulation and for supporting early seedling growth than the endosperm.

To conclude, the strong physiological correlation between seed reserves (RAP, TPS and TSS) with seed/seedling and RSA related traits corroborates our hypothesis (Chapter 2) that seedling vigour is due to a higher amount of reserves in the larger seeds and the co-location of all the RAP and TPS and several RAS and TSS QTLs with seed/seedling and RSA QTLs establishes a genetic link between seed reserves and seed quality and seedling biomass. However, a perfect co-location of all the RAP and TPS QTLs with seed/seedling and RSA QTLs suggests a strong genetic link between seed protein content and seed/seedling vigour and assigns an edge to seed protein over seed starch content for contributing to seed size and seedling establishment. This role of protein was also obvious from the tissue specific analysis where protein content of both embryo and endosperm was significantly higher and strongly correlated with phenotypic traits. However, the close quantitative but differential qualitative associations between embryo and endosperm, as well as their differential association with seedling traits, make them specific in their function. It also indicates the complexity of these seed traits.

Generalized Genetical Genomics and Metabolic-Phenotypic Associations

The recent technological advances in genome sequencing and the large scale high-throughput phenotyping technologies, such as transcriptomics, proteomics and metabolomics, have enabled plant biologists to elucidate quantitative genetic variation and associate it with phenotypes (Jansen and Nap, 2001; Keurentjes et al., 2008). Complex phenotypes, such as seed germination and seedling or plant establishment are the product of multiple genetic and environmental signals and requires the combined action of many genes. Exploring well-structured recombinant inbred lines in combination with omics analysis can help to unravel the genetic basis of such complex quantitative phenotypes. Genetical genomics is a useful approach that can effectively capture the effect of genetic

perturbation on biological systems at the molecular level. However, a molecular network also depends on environmental conditions. Therefore, a comprehensive understanding of biological systems requires studying them across multiple environments (Li et al., 2008; Ruffel et al., 2010).

Thus, to comprehensively understand how the environment interacts with the genomic encoded information (G x E), a generalized genetical genomics (GGG) approach which integrates genetic and sensibly chosen environmental perturbations has been proposed (Li et al., 2008). The integration of metabolomics with quantitative genetics is at the heart of understanding biochemical phenotypes. In Chapter 5 we used a GGG approach in which genetic perturbations of tomato seed quality were studied across multiple environments (dry and 6h imbibed seeds). Such a GGG approach was recently used to study the genetic and environmental regulation of primary metabolites in dry, 6h imbibed and germinated Arabidopsis seeds (Joosen et al., 2013). This strategy allows a crucial step toward understanding why a biological system behaves differently by varying the experimental environments. Previous reports which focused on the comparative investigation of developmental and metabolic variation propose a link between central metabolism and plant physiology, but genetic co-regulation is not frequently discussed (Keurentjes et al., 2006; Meyer et al., 2007). Using GC-TOF-MS data, the genetic regulation of variation in the tomato seed metabolome of the RIL population is described. To observe the changes in metabolite levels over the multiple environments and to identify genotype-by-environment interactions, GGG takes into account the environmental perturbation (various seed developmental stages, i.e. dry and imbibed seeds) in combination with the analysis of the genetic variation present in the RIL population. Thus it can be applied to detect the plasticity of molecular networks for seed/seedling quality phenotypes in tomato.

Interestingly, dry and 6h imbibed seeds were associated with a boost of metabolic switches that are initiated during imbibition. Abundance of different metabolites was differentially regulated by both genotype and developmental conditions. In the dry state the majority of metabolites was higher in the Pimp parent whereas in the 6h imbibed seeds the situation was completely the opposite with higher levels of both known and unknown metabolites more abundant in the MM parent. 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. On the basis of this transition, metabolites were either (1) genetically (G) regulated, irrespective of the environment, such as myoinositol, serine and asparagine, which were significantly higher in Pimp in both dry and 6h imbibed seeds and galactonic acid, GABA and monomethylphosphate in MM; (2) symmetrically environmentally regulated (E), such as most organic acids that were concomitantly down regulated and glutamine and glucose-6-phosphate that were up regulated in both parents in imbibed seed and (3) differentially environmentally regulated in different genotypes (G x E), such as urea which was

significantly higher in MM in the dry state and steeply declined in imbibed seeds, while at the same rate it increased in 6h imbibed Pimp seeds. Some metabolites such as oxalate, galactonate, phosphoric acid, threonic acid and citric acid were specifically strongly regulated by the environment rather than by the genotype. It was also important to note that where the level of most metabolites significantly decreased in 6h imbibed Pimp seeds, there was a significant increase in the level of the majority of metabolites in the MM seeds after imbibition. Thus, the higher level of most metabolites in dry Pimp seeds suggests an early maturation drying with concomitant start of mobilization of the seed reserves while the lower level of most metabolites in dry MM parent seeds might suggest delayed maturation drying of MM seed. This is also in accordance with the germination characteristics of the two parents, as Pimp germinates earlier as compared to the MM parent (Kazmi et al., 2012). Transition from reserve accumulation to maturation drying has been shown to be associated with a metabolic switch, resulting in the accumulation of distinct sugars, organic acids, nitrogen-rich amino acids, and shikimate-derived metabolites (Fait et al., 2006). 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 (Fait et al., 2006).The rapid increase in the abundance of metabolites in 6h imbibed seed of MM reveals the degradation of seed reserves and this is in accordance with the higher amount of food reserves present in the seeds of MM. It has been reported that the period of reserve accumulation in seeds is associated with a major reduction in levels of primary metabolites, including amino acids, sugars, organic acids and polyols, suggesting their utilization and incorporation into storage reserves such as protein, starch, and fatty acids (Fait et al., 2006).

The transition period from reserve accumulation to seed maturation drying is associated with a major metabolic switch, due to degradation of these seed reserves, resulting in the accumulation of distinct sugars, organic acids, nitrogen-rich amino acids, and shikimate-derived metabolites. The level of these metabolites in the seed is dependent on the rate of reserve accumulation during the seed maturation period (Fait et al., 2006). Degradation of storage proteins and its correlation with a substantial increase in the levels of free amino acids has been reported for seedlings of loblolly pine (*Pinus taeda L)* (King and Gifford, 1997). The reduction in metabolites in Pimp seeds following imbibition implies that primary metabolites are rapidly consumed to support the metabolic switch toward enhancing biosynthetic processes required for early germination. Although mobilization of oil and protein reserves occur following radical protrusion (Bewley, 1997; Eastmond and Graham, 2001), (Fait et al., 2006) have documented that active metabolic processes are already initiated during seed imbibition and that significant reduction occurs of the level of most metabolites that had accumulated during maturation drying. It seems possible that primary metabolites might be rapidly consumed to support the metabolic switch toward enhancing biosynthetic processes needed for early germination. Thus the switch from a dry

seed to a 6-hour imbibed seed is associated with a release of energy from degradation and remobilization of reserve food for seed germination and subsequent seedling growth. It was also obvious from our observations that in the 6h imbibed seeds both the number of positively correlated metabolites and the strength of correlation with seed and seedling phenotypes was significantly higher as compared to dry seed metabolites. In addition to a weak correlation and lower number of metabolites correlating with seed and seedling traits in dry seed, the percentage of negatively correlated metabolites was higher than the positively correlated metabolites in the dry state. Hence, these metabolic fluctuations can act as markers to predict the increase in the flux of specific metabolites throughout the course of tomato seed germination and seedling establishment.

The association between plant and seed primary metabolites and plant growth and biomass, as well as seed quality, has been reported in several studies (Schauer et al., 2006; Meyer et al., 2007; Prinzenberg et al., 2010; Skogerson et al., 2010; Toubiana et al., 2012). However, most of these studies could not establish a strong relationship between single metabolites and whole plant biomass (Meyer et al., 2007). More significant associations have been established between plant growth and specific combinations of metabolites (canonical correlation, CCA) (Cross et al., 2006; Lisec et al., 2009; Prinzenberg et al., 2010). In line with these findings, our current study also revealed weak to moderate correlations between single seed metabolites and seed/seedling phenotypes, although these correlations were substantially higher than those reported in previous studies between single metabolites and seedling biomass in Arabidopsis (Meyer et al., 2007). Contrarily, we observed much stronger canonical correlations, between different phenotypic traits and specific combinations of metabolites. The CCA revealed that a combination of the levels of a larger number of metabolites rather than a few individual metabolites, argued at a much closer correlation with seed mass and, indirectly, seedling growth. Thus, the present study is supported by the previous findings which reported that variation in seed mass and seedling growth might be affected by characteristic combinatorial fluctuations of metabolite levels, whereas individual metabolites may change largely independently without any major effect on plant growth (Meyer et al., 2007; Lisec et al., 2008; Sulpice et al., 2010). Investigation of the signature metabolites in CCA revealed strongly linked clusters in which both known and unknown metabolites were strongly correlated with more than one morphological trait. Known metabolites of the central metabolic pathways, including hexose sugars and hexose phosphates, such as glucose, sucrose, fructose, glucose-6-phosphate and glycerol-3-phosphate were highly represented in the CCA by correlating with several phenotypic traits. The role of these sugars in plant growth and development is well known. They play an important role in linking carbon flow from photosynthesis and starch metabolism with cell wall formation. It has been reported that a high glucose level possibly maintains the capacity of cells to divide, whereas, later in seed development, a certain sucrose level is needed to induce storage-associated cell

differentiation (Wobus and Weber, 1999). Sucrose is known to be the major transport form of carbon from source to sink tissue and thus represents the border between carbohydrate synthesis and its utilization at the whole plant level (Meyer et al., 2007). The oxidative pentose phosphate pathway also provides substrates for nucleic acid synthesis and for lignin, polyphenol and amino acid synthesis, as well as glycolysis.

Metabolites such as glycerol-3-phosphate, ethanolamine and digalactosyldiacylglycerol (DGDG) which were also highly ranked in the CCA, have been reported to play an important role in membrane/phospholipid biosynthesis (Meyer et al., 2007). These metabolites have been reported to enhance seedling growth in Arabidopsis (Kelly et al., 2003). Being a major lipid, DGDG has been documented to be required for normal growth of Arabidopsis (Awai et al., 2007) and Arabidopsis mutants deficient for this metabolite have abnormal growth (Dörmann et al., 1995).

Other major groups that were strongly represented in the CCA included the organic and amino acids. The organic acids, citrate, gluconate, malate, oxalate, galactonate, pentonic acid, phosphoric acid, quinate, threonate and salicylate were highly represented in the top ranking metabolites in the CCA in either dry or 6h imbibed seeds. Organic acid metabolism is known to be of central significance at the cellular level for several biochemical pathways, including energy production, formation of precursors for amino-acid biosynthesis and at the whole organism level in modulating adaptation to the environment (Carrari et al., 2003). It has been reported that organic acids, such as citrate, malate, and oxalate, also play a role as key components in the mechanisms that some plants use to cope with nutrient deficiencies, metal tolerance and plant–microbe interactions (Ma et al., 2001).

Among all the metabolites amino acids were the most strongly represented in the CCA in terms of both the number of correlations with seed and seedling traits as well as the strength of their correlations. These included Trp, Ser, Phe and Gln in the dry seeds and Val, Lys, Leu, Thr, Ile, Ala, Glu, Gly, Pro and Tyr in 6h imbibed seeds. Importantly, 15 of these amino acids were positively correlated with more than one physiological trait while only Asp was negatively correlated in the dry seed stage although it was also positively correlated in 6h imbibed seeds. Such a centrality of amino acids in tomato seed and fruit metabolic networks has recently also been reported by Toubiana et al.(2012). The strong representation of amino acids in the CCA reveals the significance of amino acid metabolic pathways in seedling vigour and establishment. Amino acids, such as Gln, Glu and Asp, are well known to serve as central metabolites in nitrogen assimilation in plants, whereas Ala, Ile, Leu, Met, Ser, Phe, Pro and Val are also known to be related to stress responses. Besides their role as building blocks of proteins and polypeptides, there is increasing evidence that many amino acids regulate key metabolic pathways that are necessary for plant growth, maintenance and reproduction. In general, the metabolic abundance of 6h imbibed seeds is more closely related with seed weight/size and seedling biomass than the

metabolites abundance of in dry seed..As metabolites may play a major role as nutrients for plant growth and in defense against biotic and abiotic stress, the enhancement in correlations in the imbibed stage suggests that major reserve compounds of the dry seed are degraded during imbibition and are converted to those metabolites that are needed for energy supply for germination and subsequent seedling growth. It is also obvious from ourpresent observations that seed weight/size is the major regulator of seed metabolism and this could be demonstrated from the highest number of signature metabolites mostly positively correlating with seed weight/size in the CCA followed by high CCA explained variance.

Further evidence of association between plant growth and development and primary metabolism has been demonstrated by co-location of metabolic (mQTLs) with phenotypic quantitative trait loci (phQTLs) for whole-plant biomass and yield related traits (Schauer et al., 2006; Lisec et al., 2007). Although, recently, Toubiana et al. (2012) have attempted to identify links between tomato seed metabolites and whole/mature plant phenotypes, and although extensive research exists on metabolic profiling for tomato fruit quality, no previous study exists which adequately covers the relationship between seed metabolic profiles and early seedling vigour (Schauer et al., 2006; Schauer et al., 2008). The advantage of the GGG setup is that it can detect QTLs for both the genetic as well as the genetic by environmental component in a RIL population (Joosen et al., 2013). By applying this model to the phenotypic data of the metabolic abundance in the RIL population, we were able to identify 146 mQTLs for the 166 metabolites, including 112 mQTLs for the genetic and 34 for the interaction between genetic and environmental components. The significance of the entire discipline of metabolite profiling is the concept of pathways converging to common metabolites. In many cases two or more independent pathways result in the synthesis of the same metabolite and hence, just one genetic locus may sometime not alter metabolite levels significantly and therefore may not be identified as an mQTL.

Nevertheless, in our F_8 population, we found that strongly correlated metabolites mapped to identical positions, providing evidence for the co-regulation of biologically-related pathways. Such co-location of QTLs could either be the result of closely linked genetic factors or the different metabolites may be governed by the same genetic mechanisms. Our results provide evidence of such co-regulation of biologically-related pathways. When plotted against their genomic positions, seven of such suggestive QTL clusters could be seen on 6 different chromosomes, several of which also coincided with seed and seedling trait phQTLs. In addition to co-location, strong correlations between levels of metabolites present in these QTLs hotspots was observed and many of them were also correlated with seed and seedling biomass. Amino acids were the metabolites providing the most remarkable evidence of functional clustering in both the correlation matrix and the QTL data on different chromosomes (Chapter 2 and 5) as was also shown by

Toubiana et al. (2012) when studying a tomato population of introgression lines. It has previously been suggested that clusters of co-locating QTLs are expected to correlate and the correlation may become stronger with increasing numbers of co-locating QTLs (Lisec et al., 2007). Such correlation may be the result of common genetic factors, e.g., relating to regulatory or biosynthesis genes. However, in some cases no correlation has been observed in the corresponding clusters of metabolites and phenotypic traits of the co-locating mQTLs or phQTLs (Lisec et al., 2007). This was also true for our data where in some cases we could not find correlations for the co-locating traits and in a few cases there was correlation between metabolitic or phenotypic traits but no co-location of QTLs. Such situation could be due to strong interactions of the co-locating QTLs with other loci or the metabolite may be under different metabolic and environmental controls (Lisec et al., 2007; Meyer et al., 2007).

The correlations among levels of amino acids and other metabolites, and their precursors, is in accordance with our understanding of metabolic physiology. It has been argued that the success of germination and subsequent seedling vigour are associated with the amount, degradation and mobilization of food reserves (Fait et al., 2006; Penfield et al., 2006; Bewley et al., 2012). In addition, the efficiency and mobilization of reserves during seed germination apparently also depends on the amount of reserve accumulation during seed maturation (Fait et al., 2006). The extensive genetic and physiological data sets suggest that correlation-based clustering of metabolites can be used as biomarkers for predicting seed and seedling phenotypes. Multivariate statistics has proven to be an important approach by which to predict seed germination and seedling vigour related traits. In general, therefore, it appears that integrative biomarkers have a highly significant positive or negative correlation with seed and seedling parameters and capture much of the information present in the metabolite profiles (Meyer et al., 2007; Lisec et al., 2008; Sulpice et al., 2010). However, our data also revealed that many metabolites are under the control of different genetic regulators, resulting in a distinctive genetic mapping, even within a group of strongly correlating metabolites, thus revealing the complex nature of metabolic regulation.

Post QTL Confirmation through a HIFs Approach

One of the challenges of QTL analysis is the large confidence intervals of the detected QTLs. In order to obtain more precise map information for fine mapping and cloning of casual genes, additional genome-wide analysis are required. NILs that differ only in a small genomic region at the QTL of interest provide an excellent resource for high precision QTL mapping, serving as starting material for fine mapping and cloning of gene(s) responsible for complex trait variation (Glazier et al., 2002; Balasubramanian et al., 2009). Chapter 6

deals with the development, characterization and testing of NILs present in heterogeneous inbred families (HIFs) derived from some RILs of the population to validate QTLs clusters on the bottom of chromosome 6 and 9. This approach takes benefit from the residual heterozygosity still present in the F_8 generation of our population. In this population we identified RILs that still segregated at the region of the QTLs of interest and were homozygous elsewhere and these lines were used for the construction of HIFs (Tuinstra et al., 1997; Alonso-Blanco and Koornneef, 2000; Glazier et al., 2002; Loudet et al., 2005). Using these HIFs, we were able to confirm QTLs associated with seed and seedling phenotypes at the bottom of chromosome 6 and 9. QTLs for different seed and seedling traits were mostly identified in clusters across the 12 tomato chromosomes and in particular at the bottom of chromosome 1, 6, 9, and the middle of chromosome 4. Confirmation of the QTLs on chromosome 6 and 9 regulating seed dimensions and seedling vigour related traits further strengthen our finding that the genetic variation observed in the RILs is relevant and repeatable (Chapter 2). A number of QTLs for different seed germination conditions were previously confirmed by Kazmi et al. (2013) at the same locus at chromosome 6. Thus this locus has influence on both germination characteristics and seedling growth. A number of markers associated with QTLs confirmed by HIFs at the bottom of chromosome 6 for seed germination performance as well as seedling growth and on chromosome 9 for seed and seedling traits can be used for marker-assisted selection (MAS) or help in gene cloning by fine mapping. Pleiotropic effects of these loci for different seed germination and seedling traits (Chromosome 6) and seed and seedling traits (chromosome 9) suggest that these markers can be used to breed for several seed and seedling quality aspects at once. The present study contributes to the dissection of the complex trait of seed quality and its link with seedling establishment. Further, isolation and characterization of genes and comparison of gene expression analysis at these loci can be integrated with a candidate gene approach by checking all the genes located in the interval of interest. In addition, there are still a number of other previously identified genetic regions (Chapter 2, Khan et al., 2012) that potentially influence seed size and seedling growth, which need to be confirmed with the HIF approach to allow the unequivocal determination of the regions controlling these traits.

Weaknesses, Limitations and Suggestions for Future Research

The present study demonstrates that extensive genetic variation exists in both phenotypes (germination, seed and seedling traits) and metabolite profiles. The integrated datasets from several QTL mapping (phenotypic, seed reserves, metabolic) studies have revealed the genetic changes responsible for seed quality differences and their impact on seedling characteristics among multiple RILs of the *Solanum lycopersicum* x *Solanum*

pimpinellifolium population. Furthermore, the application of high throughput metabolite profiling enables the construction of metabolic networks, which were unlikely to be elucidated from targeted small-scale approaches. As a whole, a large number of QTLs were identified for seed and seedling phenotypes, seed reserves and metabolic abundance. However, the identification of causal single nucleotide poylmorphism (QTN), regulating the trait of interest, still remains a challenge. 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 studyis that what is causing variation in the traits will also be reflected as variation at the gene expression level.

The genetical genomics approach by which we attempted to narrow down the gap between QTL and phenotype was not fully accomplished because, unfortunately, time constraints prevented us from integrating the e(xpression) QTL data in the present study. This has somewhat limited our current efforts to fully understand the genetic regulation of the mechanisms underlying complex traits by integrating genotypic, phenotypic, gene expression and metabolic data. The usefulness 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 currently in the process of analysis and will only be available for combination with the phenotypic and metabolomic data sets in the near future, to obtain a more comprehensive picture at a systems level. One other limitation of the study was that we could not elucidate the share and role of seed lipid content due to time constraint. We believe that quantification of tomato seed lipid content will provide a more complete picture of the interactions among major seed reserves (protein, starch and lipid) and their link with seed size and seedling establishment.

Potential Challenges in Genetical Characterization of Complex Traits

The exploratory findings in this study provide a new understanding of tomato seed quality at the systems level. However, a major challenge in current biology is to understand the genetic basis of variation for quantitative traits. Genes underlying quantitative traits do not necessarily affect the traits directly; they function through complex networks of transcriptional, protein and metabolic phenotypes, thus complicating the comprehension of the genetics regulating the complex traits and the downstream changes that they reflect

(Mackay et al., 2009). The power to detect and localize QTLs, and the biological context in which to place genotype-phenotype associations are the two main challenges in the genetic dissection of quantitative traits. This issue can be solved by accurate phenotypes and high density molecular genotyping to map QTLs with effect size of the magnitude we now expect, and with the high resolution and power needed to separate closely linked QTLs and to detect interactions between QTLs.

The most significant challenge is to understand the causative and correlative effects of genetic perturbations on these networks and their downstream effects on organismal phenotypes. This challenge of dissecting quantitative traits into individual genes and their causal quantitative trait nucleotides should be met in the near future by applying new sequencing and genotyping technologies to the problem, in combination with modeling to understand the effect of perturbations in one gene on the expression of other genes and the effect of those changes on the phenotypes. One of the challenges is the accurate phenotyping of both metabolites and phenotypic traits, as the status of these traits is potentially affected by the prevailing environment. Therefore they need to be phenotyped under maximally comparable conditions, which includes rigorous control of temperature, light, humidity and nutrient level at each individual plant of the population, which is a challenge that cannot be underestimated.

Accurate phenotyping may also need the availability of sophisticated and specialized equipment for fast and accurate data analysis. In the case of metabolite profiling, the main challenge is the annotation of the large number of (still) unknown metabolites. Currently, identification and classification of the numerous compounds that we obtain from advanced metabolic profiling technologies are the limiting factors in linking metabolites with phenotypic traits (Keurentjes and Sulpice, 2009). However, with the rapid advances in metabolomics and other high-throughput technologies in combination with analytical procedures such as generalized genetical genomic are now opening the way to molecular, physiological and genetic plasticity studies, triggered by environmental conditions (Jansen and Nap, 2001; Li et al., 2006; Gibson, 2008; Li et al., 2008; Joosen et al., 2013).

Thus, through a systems approach it is now possible to integrate these layers of information in order to establish biological networks that link molecular variants to organismal phenotypes (Chen et al., 2008). However, to construct meaningful networks from natural genetic perturbations, novel statistical techniques (Bayesian networks, partial correlation analysis and empirical Bayes procedures) are required (Mackay et al., 2009; Kazmi et al., 2013). Keurentjes et al. (2007) have used the power of systems genetics to unravel 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 to further investigate genome-wide gene specificity and regulation through alternative splicing and microRNAs are being developed and discussed

(Wang et al., 2008; Yin et al., 2008). In recent years, the application of massively parallel sequencing platforms for Next Generation Sequencing (NGS) methods, empowering the simultaneous sequencing of hundreds of thousands of DNA fragments, have dramatically changed the landscape and resulted insignificantly improved eQTL studies (Costa et al., 2010). This innovative platform has greatly supported the intriguing new applications such as RNA-Seq for transcriptome studies, Chip-Seq for DNA-protein interactions and CNV-Seq for large genome wide nucleotide variation analysis. Among them RNA-Seq is perhaps the most complex NGS application and a new arrival in the age of technological revolutions for accurate determination of expression levels of specific genes, differential splicing, allele-specific expression of transcripts to address many biological-related issues which are not readily achievable from previously widespread hybridization-based or tag sequence-based approaches (Wang et al., 2009; Costa et al., 2010). Nevertheless, the unprecedented level of sensitivity and the large amount of available data produced by NGS platforms provide it clear advantages. This technology empowers 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. Thus it opens up a great opportunity to study those organisms in which the genome has not yet been sequenced. Contrasting to microarrays, it does not have to cope with sequence cross-hybridization problems nor with microarray experimental designs, but allows a direct link between genotypes and phenotypes (Shendure and Ji, 2008). These technologies bring great power to make several new biological observations and discoveries. However, a great challenge remains for bioinformatics to handle NGS data (Mardis, 2008; Pettersson et al., 2009) and necessitates a considerable effort in the development of new bioinformatics tools to cope with the massive data files.

References

Alonso-Blanco C, Blankestijn-de Vries H, Hanhart C, Koornneef M (1999) Natural allelic variation at seed size loci in relation to other life history traits of *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences of the United States of America **96:** 4710

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

Awai K, Watanabe H, Benning C, Nishida I (2007) Digalactosyldiacylglycerol is required for better photosynthetic growth of Synechocystis sp. PCC6803 under phosphate limitation. Plant and Cell Physiology **48:** 1517-1523

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

Baud S, Boutin JP, Miquel M, Lepiniec L, Rochat C (2002) An integrated overview of seed development in *Arabidopsis thaliana* ecotype WS. Plant Physiology and Biochemistry **40:** 151-160

Benjamin L (1990) Variation in time of seedling emergence within populations: a feature that determines individual growth and development. Advances in Agronomy **44:** 1-25

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

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

Bewley JD, Bradford KJ, Hilhorst HWM, Nonogaki H (2012) Seeds: Physiology of Development, Germination and Dormancy. Springer Verlag

Black J (1957) The early vegetative growth of three strains of subterranean clover (*Trifolium subterrane*um L.) in relation to size of seed. Crop and Pasture Science **8:** 1-14

Carrari F, Urbanczyk-Wochniak E, Willmitzer L, Fernie AR (2003) Engineering central metabolism in crop species: learning the system. Metabolic Engineering **5:** 191-200

Chen Y, Zhu J, Lum PY, Yang X, Pinto S, MacNeil DJ, Zhang C, Lamb J, Edwards S, Sieberts SK (2008) Variations in DNA elucidate molecular networks that cause disease. Nature **452:** 429-435

Clerkx EJM, El-Lithy ME, Vierling E, Ruys GJ, Blankestijn-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

Cochrane M, Duffus C (1983) Endosperm cell number in cultivars of barley differing in grain weight. Annals of Applied Biology **102:** 177-181

Costa V, Angelini C, De Feis I, Ciccodicola A (2010) Uncovering the complexity of transcriptomes with RNA-Seq. BioMed Research International **2010**

Cross JM, Von Korff M, Altmann T, Bartzetko L, Sulpice R, Gibon Y, Palacios N, Stitt M (2006) Variation of enzyme activities and metabolite levels in 24 Arabidopsis accessions growing in carbon-limited conditions. Plant Physiology **142:** 1574-1588

Cui K, Peng S, Xing Y, Xu C, Yu S, Zhang Q (2002) Molecular dissection of seedling-vigour and associated physiological traits in rice. Theoretical and Applied Genetics **105:** 745-753

Davies DR (1975) Studies of seed development in *Pisum sativum*. Planta **124:** 297-302

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

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

Dörmann P, Hoffmann-Benning S, Balbo I, Benning C (1995) Isolation and characterization of an Arabidopsis mutant deficient in the thylakoid lipid digalactosyl diacylglycerol. Plant Cell **7:** 1801-1810

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

Eastmond PJ, Graham IA (2001) Re-examining the role of the glyoxylate cycle in oilseeds. Trends in Plant Science **6:** 72-78

Egli D, FRASER J, Leggett J, Poneleit C (1981) Control of seed growth in soya beans [*Glycine max* (L.) Merrill]. Annals of Botany **48:** 171-176

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

Evans L, Bhatt G (1977) Influence of seed size, protein content and cultivar on early seedling vigour in wheat Canadian Journal of Plant Science **57:** 929-935

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

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

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 **179:** 582-589

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

Foolad MR, Zhang LP, Subbiah P (2003) Genetics of drought tolerance during seed germination in tomato: inheritance and QTL mapping. Genome **46:** 536-545

Frary A, Fulton T, Zamir D, Tanksley S (2004) Advanced backcross QTL analysis of a *Lycopersicon esculentum* × *L. pennellii* cross and identification of possible orthologs in the Solanaceae. Theoretical and Applied Genetics **108:** 485-496

Fulton T, Beck-Bunn T, Emmatty D, Eshed Y, Lopez J, Petiard V, Uhlig J, Zamir D, Tanksley S (1997) QTL analysis of an advanced backcross of *Lycopersicon peruvianum* to the cultivated tomato and comparisons with QTLs found in other wild species. Theoretical and Applied Genetics **95:** 881-894

Fulton T, Grandillo S, Beck-Bunn T, Fridman E, Frampton A, Lopez J, Petiard V, Uhlig J, Zamir D, Tanksley S (2000) Advanced backcross QTL analysis of a *Lycopersicon esculentum* × *Lycopersicon parviflorum* cross. Theoretical and Applied Genetics **100:** 1025-1042

Gibson G (2008) The environmental contribution to gene expression profiles. Nature Reviews Genetics **9:** 575-581

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

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

Harlan JR, De Wet J, Price EG (1973) Comparative evolution of cereals. Evolution **27:** 311-325

Hicks C, Tuinstra M, Pedersen J, Dowell F, Kofoid K (2002) Genetic analysis of feed quality and seed weight of sorghum inbred lines and hybrids using analytical methods and NIRS. Euphytica **127:** 31-40

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

Hilhorst HWM, Finch-Savage WE, 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

Hodgkin T, Hegarty T (1978) Genetically determined variation in seed germination and field emergence of *Brassica oleracea*. Annals of Applied Biology **88:** 407-413

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

Jenner CF, Ugalde TD, Aspinall D (1991) The physiology of starch and protein deposition in the endosperm of wheat. Functional Plant Biology **18:** 211-226

Joosen RV, Arends D, Li Y, Willems LA, Keurentjes JJ, Ligterink W, Jansen RC, Hilhorst HW (2013) Identifying genotype-by-environment interactions in the metabolism of germinating Arabidopsis seeds using Generalized Genetical Genomics. Plant Physiology **162:** 553-566

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

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 R, Ligterink W, Hilhorst H (2013) General discussion; creating system- level models of tomato seed quality. PhD dissertation Genes for seed quality**:** 201

Kazmi R, Willems L, Ligterink W, Hilhorst H (2013) Dissection of of the complex phenotypes ofseed quality on tomato chromosomes 6 and :HIFs (Heterogeneous inbred families)confirmation. PhD dissertation Genes for seed quality**:** 179

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

Kelly AA, Froehlich JE, Dörmann P (2003) Disruption of the two digalactosyldiacylglycerol synthase genes DGD1 and DGD2 in Arabidopsis reveals the existence of an additional enzyme of galactolipid synthesis. Plant Cell **15:** 2694-2706

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 JJ, Koornneef M, Vreugdenhil D (2008) Quantitative genetics in the age of omics. Current Opinion in Plant Biology **11:** 123-128

Keurentjes JJB, Fu J, De Vos CHR, Lommen A, Hall RD, Bino RJ, van der Plas LHW, Jansen RC, Vreugdenhil D, Koornneef M (2006) The genetics of plant metabolism. Nature genetics **38:** 842-849

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 of the United States of America **104:** 1708-1713

Keurentjes JJB, Sulpice R (2009) The role of natural variation in dissecting genetic regulation of primary metabolism. Plant Signaling & Behavior **4:** 244-246

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

King JE, Gifford DJ (1997) Amino acid utilization in seeds of loblolly pine during germination and early seedling growth (I. arginine and arginase activity). Plant Physiology **113:** 1125-1135

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

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

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

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. Theoretical and Applied Genetics **110:** 742-753

Lowe L, Ries SK (1973) Endosperm protein of wheat seed as a determinant of seedling growth. Plant Physiology **51:** 57-60

Ma JF, Ryan PR, Delhaize E (2001) Aluminium tolerance in plants and the complexing role of organic acids. Trends in Plant Science **6:** 273-278

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

Mahmood T, Rahman MH, Stringam GR, Yeh F, Good A (2005) Molecular markers for yield components in *Brassica juncea* - do these assist in breeding for high seed yield? Euphytica **144:** 157-167

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

Mardis ER (2008) The impact of next-generation sequencing technology on genetics. Trends in Genetics **24:** 133

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, Törjék O, Fiehn O, Eckardt Ä, Willmitzer L, Selbig J (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

Monforte A, Tanksley S (2000) Fine mapping of a quantitative trait locus (QTL) from *Lycopersicon hirsutum* chromosome 1 affecting fruit characteristics and agronomic traits: breaking linkage among QTLs affecting different traits and dissection of heterosis for yield. Theoretical and Applied Genetics **100:** 471-479

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

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

Panthee D, Pantalone V, West D, Saxton A, Sams C (2005) Quantitative trait loci for seed protein and oil concentration, and seed size in soybean. Crop Science **45:** 2015-2022

Penfield S, Li Y, Gilday AD, Graham S, Graham IA (2006) Arabidopsis ABA INSENSITIVE4 regulates lipid mobilization in the embryo and reveals repression of seed germination by the endosperm. Plant Cell **18:** 1887-1899

Perry D (1984) Report of the vigour test committee, 1980-1983. Seed Science Technology **12:** 301-308

Pettersson E, Lundeberg J, Ahmadian A (2009) Generations of sequencing technologies. Genomics **93:** 105-111

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

Prudent M, Causse M, Génard M, Tripodi P, Grandillo S, Bertin N (2009) Genetic and physiological analysis of tomato fruit weight and composition: influence of carbon availability on QTL detection. Journal of Experimental Botany **60:** 923-937

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

Reddy VM, Daynard TB (1983) Endosperm characteristics associated with rate of grain filling and kernel size in corn. Maydica **28:** 339-355

Ries S, Everson E (1973) Protein content and seed size relationships with seedling vigour of wheat cultivars. Agronomy Journal **65:** 884-886

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

Saxena K, Faris D, Singh U, Kumar R (1987) Relationship between seed size and protein content in newly developed high protein lines of pigeonpea. Plant Foods for Human Nutrition (Formerly Qualitas Plantarum) **36:** 335-340

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 (2006) Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. Nature Biotechnology **24:** 447-454

Schauer N, Zamir D, Fernie AR (2005) Metabolic profiling of leaves and fruit of wild species tomato: a survey of the *Solanum lycopersicum* complex. Journal of Experimental Botany **56:** 297-307

Shendure J, Ji H (2008) Next-generation DNA sequencing. Nature Biotechnology **26:** 1135-1145

Sheoran IS, Olson DJH, Ross ARS, Sawhney VK (2005) Proteome analysis of embryo and endosperm from germinating tomato seeds. Proteomics **5:** 3752-3764

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

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

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. Plant Cell **22:** 2872-2893

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

Thomas J, Prasad P, Boote K, Allen Jr L (2009) Seed composition, seedling emergence and early seedling vigour of red kidney bean seed produced at elevated temperature and carbon dioxide. Journal of Agronomy and Crop Science **195:** 148-156

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. 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

Weis IM (1982) The effects of propagule size on germination and seedling growth in *Mirabilis hirsuta*. Canadian Journal of Botany **60:** 1868-1874

Weschke W, Panitz R, Sauer N, Wang Q, Neubohn B, Weber H, Wobus U (2000) Sucrose transport into barley seeds: molecular characterization of two transporters and implications for seed development and starch accumulation. Plant Journal **21:** 455-467

Whittington W (1973) Genetic regulation of germination. Seed ecology: W. Hyedecter, Butterworth, London: 5-30

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

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

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. Proceedings of the National Academy of Sciences of the United States of America **96:** 6529-6534

Zhang J, Maun M (1993) Components of seed mass and their relationships to seedling size in *Calamovilfa longifolia*. Canadian Journal of Botany **71:** 551-557

Summary

The main objective of this thesis was to find the genetic and physiological components governing seed quality traits in terms of the seed's ability to germinate and survive as seedling until it is autotrophically growing under both normal and stress conditions. In the past, plant breeding has mainly focused on crop yield and disease resistance. However, it has now become equally essential to produce seeds that germinate rapidly and uniformly and establish vigourous seedlings under various environments..

Seed quality is determined by genetic and physiological factors and their interaction with the environment. The significant role of seed characteristics, such as seed weight, size, and purity in seed germination and seedling vigour has been demonstrated in several studies. However, at a systems level there is little or no information available about the genetic components of seed and seedling quality traits and their mutual interactions in tomato. To explore the genetic, physiological and environmental aspects underlying seed quality in tomato, an integrated study combining physiology, genetics and genomics was carried out to provide a basis for the improvement of the percentage, rate and uniformity of germination and early seedling development, and for an increase of the range of environmental conditions for germination. Our present study is, to the best of our knowledge, the first systems analysis of the genetics of seed and seedling traits, which will potentially add to what is known about tomato seed quality.

Plants cannot escape from the environment at which the seed is dispersed. Hence, to survive under diverse environments and to empower production, plants have to adapt to such environments, which results in considerable genetic variation. This genetic variation is a great resource to explore the mechanisms of adaption. We proceeded with exploring the natural variation present in a recombinant inbred line population (RIL) of *Solanum lycopersicum* x *Solanum pimpinellifolium* to elucidate the molecular-genetic mechanisms regulating seed quality. The two parents of the RIL population have considerable genetic and physiological variation in terms of germination, seed and seedling morphology and resistance to stress conditions, whereas the RILs have intermediate characteristics for many traits. This RIL population is therefore particularly suitable to locate the genomic regions with genetic differences that influence seed size, germination and seedling traits. Such regions are commonly known as quantitative trait loci (QTLs).

In Chapter 2 we analyzed the RIL population for genetic and physiological interactions between seed dimensions and early seedling vigour. Good seedling establishment and seedling vigour are essential for sustainable and profitable crop production and are considered to be the most critical stage of a developing crop. We identified numerous small and large effect QTLs along the tomato genome in the *S. lycopersicum* x *S. pimpinellifolium* RIL population governing seed, seedling and root system architecture traits. These results established a strong genetic relationship between various

seed dimension and seedling vigour related traits with mostly the same direction of the genetic effects. In addition, the genetic relationship was also followed by strong physiological associations, as strong correlations were observed among the phenotypic values for seed and seedling traits. This indicates that seed size and -weight have a strong effect on seedling growth and, ultimately, affect the initial growth of the main root and the upward growth of the shoot. Interestingly, the effect of seed weight or size was more pronounced on seedling growth under nutrient stress. These results are of great significance for the isolation of the corresponding genes and elucidation of the underlying genetic and physiological mechanisms controlling these traits. Another important finding was that germination performance and seed traits are controlled by different independent genetic loci. This was obvious from our results, as we could not establish any relationship between seed traits and seed performance, such as maximum germination (%), rate of germination or uniformity of germination. Whereas plant processes are frequently controlled by the action of several small-to large-effect genes that follow classic Mendelian inheritance, our results suggest that seed quality is quantitatively and genetically complex. Seed quality and seed performance are closely related and linked characteristics but are not the same aspect of seeds. The term seed quality has a very broad meaning that includes essentially 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, vigour, 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.

Germination characteristics are some of the aspects that determine seed performance. To elucidate the performance of the seed, we investigated the germination potential of the RIL population under both normal and stress conditions in Chapter 3. Our study revealed 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. Thus, QTL mapping with SNP-based linkage maps provides a robust assessment of the genetic architecture of the tomato genome in terms of the magnitude of QTL effects, QTL-environment interactions and putative pleiotropy. Obviously, this approach could help in the simultaneous improvement of seed traits and progress towards identifying the genes controlling seed quality and performance.

In addition to other attributes, seed quality also depends on the amount, as well as composition of protein, starch and oil in the seed. These reserves are functionally dependent on the C-N balance, central metabolism and sink-source interactions during

development on the mother plant and are frequently referred to as complex traits. The major significance of these reserves is that they support early seedling growth when degraded upon germination and, therefore, contribute to crop establishment. Therefore, the control of establishment and vigour of the young seedlings may be estimated by the quality of the seed, its response to diverse environments and the food reserves, which are available to sustain the seedling until it becomes an independent, autotrophic organism, able to use light energy. Large seeds produce vigorous seedlings due to a higher amount of reserves. Seed weight in different crops is often positively correlated with the amounts of reserves of the seed and, in turn, seed reserves are frequently positively correlated with seedling growth and biomass. To find the genetic and physiological links between seed reserves and seed and seedling quality traits, Chapter 4 explored the natural variation present in the RIL population for seed reserves by integrating QTL mapping and correlation analysis. Our study revealed that a diverse set of QTLs for seed protein and starch content map to the 12 chromosomes of tomato, influencing reserve food phenotypes. Importantly, our results demonstrate an overlap of most seed protein and several starch related QTLs with numerous seed and seedling traits, with the same direction of the genetic effect. This suggests a close genetic link between seed reserves and seed and seedling biomass. The phenotypic values for both abundance and total amount of protein, as well as total amount starch per seed reveal that larger seeds have significantly higher amounts of these seed reserves as compared to small seeds. In addition, we also show that the total amount of protein in tomato seeds is significantly higher than the starch content. The Pearson correlation analysis between the phenotypic values of the seed reserves and seed and seedling phenotypes indicates a strong physiological association between seed weight and size and seedling growth. In the same chapter we also analyzed the contribution of seed tissues to seed weight and observed that both endosperm and seed coat account for 60-70% and the embryo 30-40% of tomato seed weight, whereas the embryo accounts for about 80% of the seed size. In addition to their contribution to seed weight and size, our results also demonstrated variation in accumulation of seed reserves in the different seed tissues with significantly higher levels of protein in the embryo as compared to the endosperm. Remarkably, the embryo traits and embryo protein content generally revealed stronger correlation with seed and seedling phenotypes than the endosperm.

Although previous studies relating to seed quality phenotypes appear to be promising, none of these combined genetic, phenotypic and metabolic datasets to unravel seed quality, with respect to seedling growth in tomato. In Chapter 5 we integrated such datasets which contributed to a comprehensive biological understanding of observed phenotypic and metabolic differences between the RILs. The metabolic profiling of tomato seeds revealed great genetic variation at the level of the various metabolites in the seeds of the two parents, as well as of the RIL population. We applied a generalized genetical genomics model for mapping metabolic quantitative trait loci (mQTLs). The advantage of

this model is that it incorporates genetic, as well as environmental effects. It takes into account chosen environmental perturbations (here different seed developmental stages, i.e. dry and 6h imbibed seeds) in combination with the genetic variation present in the RIL population, to understand the change of metabolites over the different environments and to detect genotype-by-environment interactions. By applying this approach we were able to map genetic mQTLs and mQTLs that are the result of interaction between the genetics and environmental perturbations (Genetic x Environment QTLs). QTL analysis of 160 detected metabolites in the RIL population revealed QTL hotspots for 60% known and 40% unknown metabolites. Among all of the 112 mQTLs, 70% were detected for the genetic component while 30% were related to genetic-by-environment interactions (G x E QTLs). Several groups of these mQTLs co-located with seed weight and size and seedling biomass QTLs, reinforcing the concept that metabolite profiles of seeds and seedling phenotypes are genetically linked. The transition from dry to 6h imbibed seeds was associated with programmed metabolic switches, exhibiting various metabolites that were synthesized in accordance with demand and possible utilization in seed germination and early seedling growth. The metabolites most relevant to seed and seedling traits were extracted using multivariate statistics (canonical correlation). A strong canonical correlation revealed that a specific combination of metabolites could explain the phenotype of seed and seedling. Particularly, the metabolites of 6h-imbibed seeds explained positively and strongly the phenotype of seed and seedling, as compared to those of the dry seeds where low and mostly negative correlations were observed. This suggests that metabolic pathways proceeding during the 6h of imbibition have to synthesize certain metabolites for the demand of seed germination and seedling growth. This finding may contribute to enhancing our understanding of the role of highly correlated metabolites in building a comprehensive picture of associations between metabolites and seed and seedling phenotypes and provide a first step to the unravelling of the complex metabolic networks that influence seed and seedling quality.

Confirmation and fine mapping of QTLs responsible for complex trait variation can be achieved by the development of near isogenic lines (NILs). In this study, these NILs were derived from the RILs having residual heterozygosity left in the F_8 generation. These RILs were genotyped to find lines that still segregated only at the region around the QTL of interest. These types of NILs are called heterogeneous inbred families (HIFs) (Chapter 6). One QTL was confirmed on the bottom of chromosome 6 using HIFs affecting hypocotyl length. Several seed quality QTLs were confirmed using HIFs affecting seed traits on chromosome 9. This confirmation of QTLs with respect to the variation observed in hypocotyl length and seed dimensions was relevant to the repeatability of the variation observed in our previous phenotypic experiments with the RIL population. These genomic regions controlling seed and seedling traits are likely candidates for further fine mapping,

isolation and characterization of the causal genes and the molecular characterization of the pathways in which they are involved.

This thesis provides an illustration of the association between seed quality and seed germination and seedling phenotypes through integrating genotyping, phenotyping, seed reserve contents and molecular phenotyping in generating a novel understanding of seed phenotypes and their interaction with the environment. The integration of all these datasets has facilitated the identification of novel biomarkers that could be used in further seed quality testing by integration of additional analysis such as transcriptomics. Thus, integration of the current findings with multiple biological information from DNA and gene expression studies with the phenotype and environment may elucidate the relationship among key genetic factors in distinct genomic regions and their effect on seed and seedling phenotypes.

Samenvatting

Het belangrijkste doel van dit proefschrift was het vinden van de genetische en fysiologische componenten die bepalend zijn voor zaadkwaliteit in termen van het vermogen tot kieming van het zaad en overleven als zaailing tot deze autotroof kan groeien onder zowel normale als stress omstandigheden. In het verleden heeft de plantenveredeling zich voornamelijk gericht op opbrengst en ziekteresistentie van het gewas. Het is nu echter even essentieel geworden om zaden te produceren die snel en uniform kiemen en krachtig groeiende zaailingen voortbrengen die ook gedijen onder minder optimale omstandigheden.

Zaadkwaliteit wordt bepaald door zowel genetische als fysiologische factoren en hun interactie met de omgeving. Een significante rol van zaaddimensies zoals gewicht, afmetingen en zuiverheid in zaadkieming en zaailingvigour is in veel studies aangetoond. Er is echter op systeemniveau nauwelijks informatie beschikbaar over genetische componenten van kwaliteitseigenschappen van zaad en zaailing, en hun onderlinge interacties, in tomaat. Ten einde de genetische, fysiologische en omgevingseigenschappen van zaadkwaliteit in tomaat te verkennen werd een geïntegreerd onderzoek uitgevoerd, waarin fysiologie, genetica en genomica werden gecombineerd, om een basis te vinden voor verbetering van kiemingspercentage en -uniformiteit alsmede vroege ontwikkeling van de zaailing, onder een reeks van omgevingscondities. Onze studie verschaft, voor zover wij weten, de eerste systematische analyse van de genetica van zaad- en zaailingeigenschappen die een mogelijke bron vormen voor een betere begripsvorming van zaadkwaliteit in tomaat.

Planten kunnen niet ontsnappen aan de omgeving waarin het zaad is verspreid. Om toch te kunnen overleven onder diverse omgevingsomstandigheden en productief te zijn, moeten planten zich aanpassen aan de verschillende omstandigheden wat resulteert in aanzienlijke genetische variatie. Deze genetische variatie vormt een onuitputtelijke bron voor het bestuderen van adaptatiemechanismen. Wij hebben gebruik gemaakt van de natuurlijke variatie die aanwezig is in een populatie van recombinante inteeltlijnen (RILs) van de ouders *Solanum lycopersicum* x *Solanum pimpinellifolium* ten einde de moleculair-genetische mechanismen te ontrafelen die zaadkwaliteit reguleren. De beide ouders van de RIL populatie vertonen aanzienlijke genetische en fysiologische variatie in termen van kieming, zaad- en zaailingmorfologie en resistentie tegen stress condities, terwijl de RILs tussenliggende waarden laten zien voor de vele eigenschappen. Deze RIL populatie is daarom vooral geschikt om de genomische gebieden te lokaliseren waarin genetische verschillen zaadgrootte, kieming en zaailingeigenschappen beïnvloeden. Zulke gebieden worden Quantitative Trait Loci (QTLs) genoemd..

In Hoofdstuk 2 hebben wij deze RIL populatie geanalyseerd voor genetische en fysiologische interacties tussen zaaddimensies en zaailingvigour. Goede zetting en vigour

van zaailingen zijn van groot belang voor duurzame en winstgevende gewasproductie en vertegenwoordigen de meest kritieke fase van een beginnend gewas. Wij hebben talloze QTLs gevonden met zowel grote als kleine effecten over het gehele tomatengenoom in de *S. lycopersicum* x *S. pimpinellifolium* RIL populatie voor eigenschappen van zaad en zaailing grote en wortelarchitectuur. Deze resultaten hebben een sterke genetische relatie vastgesteld tussen de verschillende zaaddimensies en eigenschappen gerelateerd aan zaailingvigour met meestal dezelfde richting van de genetische effecten. Daarnaast gingen deze genetische relaties gepaard met sterke fysiologische associaties in de vorm van sterke correlaties tussen fenotypische waarden voor zaad- en zaailingeigenschappen. Dit geeft aan dat zaadgrootte en -gewicht een sterk effect hebben op zaailinggroei en, uiteindelijk, de initiële groei van de hoofdwortel en de opwaartse groei van de stengel. Interessant is dat het effect van zaadgewicht of -grootte meer uitgesproken was onder condities met nutriëntenstress. Deze resultaten zijn van groot belang voor de isolatie van de corresponderende genen en opheldering van de onderliggende genetische en fysiologische mechanismen die deze eigenschappen bepalen. Een andere belangrijke vondst was dat kiemingsprestaties en zaadeigenschappen bepaald worden door verschillende onafhankelijke genoomregio's. Dit was evident vanwege het ontbreken van elke relatie tussen zaaddimensies en zaadprestaties zoals maximale kieming (%), kiemsnelheid en -uniformiteit. Terwijl processen in planten vaak gestuurd worden door de werking van verscheidene genen met grote of kleine effecten die de klassieke Mendeliaanse overerving volgen, wijzen onze resultaten erop dat zaadkwaliteit zowel kwantitatief als genetisch complex is. Zaadkwaliteit en zaadprestaties zijn nauw verwante en verbonden karakteristieken, maar vertegenwoordigen verschillende aspecten van het zaad. De term 'zaadkwaliteit' heeft een zeer brede betekenis waaronder alle genetische, fysiologische en fysische kenmerken van het zaad vallen. Er kan worden beweerd dat sommige zaadkenmerken, zoals fysische zuiverheid, vigour, kiemkracht, snelle kieming en productie van normale zaailingen onder verschillende (stress) omstandigheden van groot belang zijn voor gewasproductie, de grenzen van kwaliteit overschrijden en daarom het best beschreven en begrepen kunnen worden in termen van 'prestatie'.

Kiemingskenmerken bepalen deels de zaadprestatie. Teneinde aspecten van zaadprestatie op te helderen hebben we in Hoofdstuk 3 het kiemingspotentiaal van de RIL populatie onderzocht onder optimale condities en onder omgevingsstress. Het bleek dat een diverse set van fenotypische kenmerken gekarteerd kon worden op de twaalf chromosomen van het tomatengenoom die van invloed waren op zaadkwaliteitsfenotypen onder optimale condities en onder condities van zout-, osmotische-, koude-, hoge temperatuur- en oxidatieve stress. De QTL benadering bleek waardevol, niet alleen voor het ophelderen van de genetica, maar ook van de fysiologische achtergrond van de zaadkwaliteitsfenotypes. Er zijn QTLs gevonden die specifiek waren voor stress of juist voor de afwezigheid van stress en die bepalend waren voor het kiemingsproces onder de

verschillende omgevingsomstandigheden. Daarom verschaft QTL kartering met op SNPs gebaseerde genetische kaarten een robuuste bepaling van de genetische architectuur van het tomatengenoom in termen van de grootte van de QTL effecten, interacties tussen QTLs en omgeving en mogelijke pleiotropie. Het is duidelijk dat deze benadering kan helpen bij de gelijktijdige verbetering van verschillende zaadkwaliteitskenmerken en bij het identificeren van de genen die bepalend zijn voor deze kenmerken.

Naast genoemde kenmerken is zaadkwaliteit ook afhankelijk van de hoeveelheid en samenstelling van eiwit, zetmeel en olie in het zaad. Deze voedselreserves zijn afhankelijk van de C-N balans, het centrale metabolisme alsmede sink-source interacties tijdens ontwikkeling aan de moederplant en worden meestal geclassificeerd als 'complexe kenmerken'. De belangrijkste rol van deze reserves is het ondersteunen van vroege groei van de zaailing als deze worden afgebroken tijdens het kiemingsproces en daardoor bijdragen aan zetting van het gewas. Dus, samengevat, zijn zetting en vigour van de jonge zaailing afhankelijk van de zaadkwaliteit, de reactie op de omgeving en de voedselreserves die beschikbaar zijn om de zaailing te onderhouden totdat deze uitgegroeid is tot een onafhankelijk en autotroof organisme dat lichtenergie kan gebruiken.

Grote zaden produceren zaailingen met hoge vigour vanwege de grotere hoeveelheden van voedselreserves. In diverse gewassen is het zaadgewicht vaak positief gecorreleerd aan de hoeveelheid reservevoedsel en zijn zaadreserves vaak gecorreleerd aan zaailinggroei en biomassa. Teneinde de genetische en fysiologische verbanden te vinden tussen de reserves en kenmerken van zaad- en zaailingkwaliteit, wordt in Hoofdstuk 4 de natuurlijke variatie in de RIL populatie onderzocht voor zaadreserves middels integratie van QTL kartering en correlatieanalyse. Ons onderzoek identificeerde een diverse set van QTLs voor eiwit- en zetmeelgehalte op de twaalf chromosomen van tomaat die fenotypes voor reservevoedsel bepaalden. Belangrijk is dat onze resultaten wijzen op een overlap van de meeste zaadeiwit- en verscheidene zetmeel gerelateerde QTLs met vele zaad- en zaailingkenmerken met een zelfde richting van het genetische effect. Dit suggereert een nauw verband tussen zaadreserves en biomassa van zaailingen. De fenotypische waarden van zowel het gehalte als de totale hoeveelheid eiwit, evenals de totale hoeveelheid zetmeel per zaad wijzen erop dat grotere zaden significant grotere hoeveelheden van deze reserves bezitten, vergeleken met kleinere zaden. Daarnaast laten we ook zien dat de totale hoeveelheid aan eiwit in tomatenzaad significant groter is dan de hoeveelheid zetmeel. De analyse van de Pearson correlaties tussen fenotypische waarden van de zaadreserves en zaad- en zaailingfenotypen suggereren een sterke fysiologische associatie tussen zaadgewicht, -grootte en zaailinggroei.

In hetzelfde hoofdstuk hebben we ook de bijdrage van de afzonderlijke zaadweefsels aan het zaadgewicht geanalyseerd en hebben waargenomen dat endosperm en zaadhuid samen voor 60-70% van het zaadgewicht verantwoordelijk zijn en het embryo 30 tot 40%, terwijl het embryo voor ongeveer 80% de zaadgrootte bepaalt. Naast hun

bijdrage aan zaadgewicht en -grootte laten onze resultaten ook zien dat er variatie is in de opslag van de reserves in de verschillende zaadweefsels met significant grotere hoeveelheden eiwit in het embryo vergeleken met het endosperm. Het is opmerkelijk dat embryokenmerken en eiwitgehalte van het embryo in het algemeen sterker correleerden met zaad en zaailingfenotypes dan van het endosperm.

Hoewel eerdere studies over zaadkwaliteitsfenotypes veelbelovend lijken, is er niet één die genetische, fenotypische en metabolische datasets heeft gecombineerd om zaadkwaliteit te ontrafelen in relatie met zaailinggroei in tomaat. In Hoofdstuk 5 hebben wij dergelijke datasets geïntegreerd als bijdrage aan een compleet biologisch begrip van waargenomen fenotypische en metabolische verschillen tussen de RILs. Het metabolisch profileren van tomatenzaad ontsluierde grote genetische variatie van de metabolietniveau's in de zaden van beide ouders en van de RILs. Wij hebben een 'generalized genetical genomics' model toegepast voor de kartering van metabolische QTLs (mQTLs). Het voordeel van dit model is dat het zowel genetische effecten als die van de omgeving incorporeert. Rekening houdend met gekozen perturbaties van de omgeving (hier verschillende stadia van zaadontwikkeling, nl. droge en 6-uur geïmbibeerde zaden) en gecombineerd met de genetische variatie van de RIL populatie kan het model inzicht verschaffen in de veranderingen van de metabolieten in de verschillende omgevingen alsmede interacties tussen genotype en omgeving.

Met deze benadering waren wij in staat om genetische mQTLs en mQTLs die het gevolg zijn van de interactie tussen de genetica en omgevingsperturbaties (G x E mQTLs) te karteren. QTL analyse van 160 gedetecteerde metabolieten in de RIL populatie resulteerde in QTL hotspots voor 60% van de bekende en 40% van de onbekende metabolieten. Van alle 112 mQTLs, was 70% gerelateerd aan de genetische component en 30% aan de G x E component. Verscheidene groepen van deze mQTLs co-loceerden met QTLs voor zaadgewicht en -grootte en met zaailing biomassa, wat het idee versterkt van een genetisch verband tussen metabolietprofielen van zaad- en zaailingfenotypes. De overgang van droog naar 6-uur geïmbibeerd zaad werd gekarakteriseerd door geprogrammeerde metabolische 'schakelingen' die verschillende metabolieten lieten zien die werden gesynthetiseerd naar vraag en mogelijke verwerking in zaadkieming en zaailinggroei. De metabolieten die het meest relevant waren voor zaad- en zaailingkenmerken werden vastgesteld middels 'multivariate statistics' ('canonical correlation'). Een sterke 'canonical correlation' geeft aan dat een specifieke combinatie van metabolieten een zaad- of zaailingfenotype kan verklaren. Met name de metabolieten van 6-uur geïmbibeerde zaden konden de zaad- en zaailingfenotypen positief en stellig verklaren in tegenstelling tot die in droge zaden waar meestal lage en negatieve correlaties werden waargenomen. Dit suggereert dat metabolische reactieroutes die tijdens de 6-uur imbibitie worden gevolgd moeten leiden tot de synthese van bepaalde metabolieten die gevraagd worden voor zaadkieming en groei van de zaailing. Dit resultaat kan bijdragen aan vergroting van ons

begrip van de rol van hoog-gecorreleerde metabolieten in een allesomvattend beeld van associaties van metabolieten met zaad- en zaailingfenotypes en kan een eerste stap zijn naar ontrafeling van complexe metabolische netwerken die zaad- en zaailingkwaliteit beïnvloeden.

Bevestiging en 'fine-mapping' van QTLs die verantwoordelijk zijn voor complexe variatie in zaadkenmerken kunnen worden gedaan met behulp van de ontwikkeling van 'near isogenic lines' (NILs). In onze studie werden deze NILs afgeleid van de RILs die een resterende heterozygotie hadden in de F_8 generatie. Deze RILs werden gegenotypeerd om lijnen te vinden die nog segregeerden in het gebied rond de betreffende QTLs. Deze NILs worden heterogene inteelt families (HIFs) genoemd (Hoofdstuk 6). Met behulp van HIFs werd eén QTL voor hypocotyl lengte op de bodem van chromosoom 6 bevestigd. Verschillende QTLs voor zaadkwaliteit op chromosoom 9 werden ook bevestigd met HIFs. Deze bevestiging van QTLs in relatie met de waargenomen variatie in hypocotyllengte en zaaddimensies was relevant voor de herhaalbaarheid van de variatie die was waargenomen in eerdere experimenten met de RIL populaties. Deze genomische gebieden die zaad- en zaailingeigenschappen bepalen zijn mogelijke kandidaten voor verdere kartering, isolatie en karakterisering van de oorzakelijke genen alsmede de moleculaire ontrafeling van de reactieroutes waarbij deze zijn betrokken.

Dit proefschrift verschaft voorbeelden van associaties tussen zaadkwaliteit en fenotypes van zaadkieming en zaailing door integratie van genotypering, fenotypering, gehalte aan zaadreserves en moleculaire fenotypering en genereert zo een nieuw begrip van zaadfenotypes en hun interacties met de omgeving. De integratie van alle datasets heeft de identificatie mogelijk gemaakt van nieuwe biomerkers die gebruikt kunnen worden in het testen van zaadkwaliteit, ook in combinatie met additionele analyse, zoals transcriptomics. Integratie van onze resultaten met veelzijdige biologische informatie uit DNA en genexpressie studies en met zowel fenotype als omgeving kan de relatie verklaren tussen genetische sleutelfactoren in bepaalde genomische gebieden en hun effect op zaad- en zaailingfenotypes.

Acknowledgements

After completing my MSc (Hons) from a local university, it was my dream to obtain a PhD degree from a well-known and top-ranking university abroad, which turned into a reality in 2008 due to a scholarship award from Higher Education Commission (HEC) Government of Pakistan and after I received the acceptance letter from my supervisor. Now I am feeling proud to be a PhD graduate from Wageningen UR which ranks second in the field of agriculture according to a world ranking in 2013.

The successful completion of my thesis and subsequent PhD degree has been a long, fruitful and learning journey. I was lucky to be accepted for my PhD studies in the Wageningen Seed Lab at the department of Plant Physiology. I found people in this department that have expertise in multiple scientific disciplines and that were always available to help, support and guide me at each stage of my studies and experiments. Without their help and support I would not have been able to have my thesis finished as I have now. Therefore, they deserve my appreciation and acknowledgement for their contribution to my thesis. I am absolutely delighted to take this opportunity to thank them all.

As it has been said that the first impression is the last impression and it became known to me from my first telephonic conversation from Pakistan with **Henk Hilhorst** that I have been accepted for my PhD studies by a right person and I was coming to the right place to pursue my PhD studies. Henk Hilhorst; the generous, the caring, the genius, always supported me at each stage of my PhD studies. He never allowed any technical or financial problem to challenge me during my study. He constantly and convincingly conveyed a spirit of adventure in regard to research and brought unique perspectives. He contributed enormously to the success of my PhD thesis. It is my pleasure to express my heartiest gratitude for all his contributions of time, ideas, and supervision to make my Ph.D. experience productive and stimulating.

Wilco Ligterink, the all-rounder in the field of science; the one who makes himself happy by helping others, and the joy and passion he has, was always enthusiastic and motivational for me. He never made me feel aware of his eventful schedule of managing multiple tasks and was always available to support and guide me. His contribution to my thesis is invaluable. I feel happy to express my deepest appreciation to my supervisor **Wilco Ligterink** for his generosity with time, advice and supervision. Without his guidance and persistent help and support this dissertation would not have been possible.

My promoter professor Harro Bouwmeester persistently encouraged and supported me throughout of my PhD studies. I am especially thankful for his valuable suggestions and guidance.

Leo Willems, the pivotal member of WSL, the energetic, the life time performer. Although quite younger than me, I have a place for him along with my other teachers. He not only guided me in technical skills, but also contributed to every aspect of my PhD thesis. I feel proud to express my heartiest thanks for his contribution to the success of my PhD thesis. I appreciate the way he managed experiments, the time he spared and the efforts he did to plan and perform experiments related to my thesis.

I consecrate my sincere thanks to Jamar Diaan for conveying her technical support in the analysis of tomato seed starch content. I gratefully acknowledge Maarteen Koorneef, Bas Dekkers and Leonie Bentsink, who always gave me suggestions and advised me whenever I needed their help.

Rashid Kazmi and Ronny Joosen, the fellow pioneer PhDs of WSL, as we worked together during the whole period of our PhD studies and I offer my gratitude and very much appreciate their enthusiasm, passion, and willingness to help and to cooperate with each other in the time of necessity.

The members of Plant Physiology and especially Wageningen Seed Lab have contributed tremendously to my personal and professional time at Wageningen. The group has been a source of friendships as well as good advice and collaboration. Other past and present group members that I have had the pleasure to work with or alongside are the numerous guest and rotation students who have come through the WSL. Here it would be injustice to not acknowledge the support and cooperation by Rina Anthonijsz, the secretary of Plant Physiology throughout my studies. I am very grateful to her for her cooperation and academic guidance.

I give my deepest thanks to my MSc supervisor, Prof. Dr. Amanullah Jan at the Department of Agronomy, KPK Agricultural University Peshawar, Pakistan for his kind support and valuable guidance during my preparations for the PhD studies.

For me, being abroad was a new and challenging experience in a different culture and environment, but it all turned into a memorable and nice stay when my friendship with non-Pakistanis was intermingled with those of Pakistani fellows. I had a very good and memorable time with all my Pakistani friends in Wageningen and elsewhere in the Netherlands. I extend my heartfelt thanks to Dr. Nazir Ahmed Khan, Dr. Sabaz Ali Khan, Dr. Sultan Mahmood, Dr. Muhammad Sohail Khan, Dr. Zeshan Hassan, Dr. Muhammad Jamil, Imran Haider, Mazhar Ali Khan, Mubarak Ali, Masood Awan, Dr. Ghulam Mustafa Shah, Dr.

Maan Abid, Dr. Muhammad Imtiaz Rashid, Dr. Ghulam Abbas Shah, Dr. Kashif Iqbal Khan, Zakir Hussain Dahri, Shafqat Qaisrani, Munawar Shah, Shahid Iqbal, Dr Sajid Rehman and Dr. Nadeem Khan.

I gratefully acknowledge the financial, academic and technical support from higher education commission (HEC) of Pakistan and Technology foundation STW-NWO of The Netherlands. I am equally grateful to The Netherlands Organisation for International Cooperation in higher education (NUFFIC) for their excellent support, assistance and cooperation during the whole period of my study in The Netherlands.

Finally, I express my deepest gratitude to my family: my brother Abdullah Khan, my nephews Noor Shah Khan, Zainullah, Ahmmadullah, Asfaqullah, khalilur-Rehman, Rehmanullah and my sons Inamullah and Abid and my cousins Dr. Rehan, Dr. Dost Muhammad and my friend and ex-colleagues Khalid Naved, Shahid Mahmood and Wiqar Ahmad for their love, support and encouragement. They always have and provided unwavering love and encouragement throughout my stay in The Netherlands. I dedicate my thesis to my most affectionate and kind brother Haji Abdullah Khan as a small token of appreciation for everything that he has done for me.

Curriculum Vitae

Noorullah Khan was born in a small town Allam of North Waziristan Agency of Pakistan with the most beautiful mountains and green forests of Chagdara Barmal, Razmak, Shawal, and surrounded by Kuram River to the North and Gomal River to the South and Tochi river in the centre which always attracts the people from the plain area of Pakistan, especially in summer.

Education & Qualifications

Sep 2008-Sep 2013:	Wageningen Seed Lab, Lab of Plant Physiology, Wageningen UR, The Netherlands: PhD in Plant Physiology (Seed Biology). Thesis title: Genetic and Physiological Quality of Tomato Seed and Seedlings. Supervisors: Dr. Henk W.M. Hilhorst and Dr. Wilco Ligterink
1998-2001:	Khyber Pakhtunkhwa Agricultural University Peshawar Pakistan: MSc in Agronomy
1994-1998:	Faculty of Agriculture Gomal University Dera Ismail Khan Pakistan: BSc (Hons) in Agronomy

Research and Work Experience

Aug 2008-Sep 2008:	Pakistan Agricultural Research council (PARC) Islamabad, Pakistan. Promoted to Scientific officer (SPS 8/PBS 17). Posted at the Agricultural Research Institute (ARI), Dera Ismail Khan, Khyber, Pakhtunkhwa province, Pakistan, in one of the Coordinated Research Programs of PARC on sugar crops such as sugarcane and sugar beet. Focusing on evaluation of sugarcane and sugar beet varieties.
Apr 2005-Aug 2008:	Pakistan Agricultural Research council (PARC) Islamabad, Pakistan. Assistant Scientific Officer (SPS 7/PBS 16). Posted at Agricultural Research Institute (ARI), Dera Ismail Khan, Khyber, Pakhtunkhwa province, Pakistan, in one of the Coordinated Research Programs of PARC on sugar crops such as sugarcane

	and sugar beet. Focusing on evaluation of sugarcane and sugar beet varieties.
Jul 2004-Apr 2005:	Pakistan Agricultural Research council (PARC) Islamabad, Pakistan. Research Fellow. Maize, sorghum and millet program at the National Agricultural Research Centre (NARC), Islamabad, Pakistan. Field crop research with focus on evaluation of maize, sorghum and millet varieties.
Jul 2003-Jul 2004:	Pakistan Agricultural Research council (PARC) Islamabad, Pakistan, Trainee Scientist, Maize, sorghum and millet program at the National Agricultural Research Centre (NARC), Islamabad, Pakistan. Field crop research with focus on evaluation of maize, sorghum and millet varieties.
Jan 2003-Jul 2003:	Islamic Relief, an international NGO at Kuttan AJK Pakistan, Agricultural Specialist. Research and extension work on field crops and orchards

Publications

Related to this Thesis

- **Khan N.**, Kazmi R.H., Willems L.A.J., van Heusden A.W., Ligterink W., Hilhorst H.W.M. (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.**, 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.**, Willems L.A.J., Ligterink W., Hilhorst H.W.M. (2013) Genetic Analysis of Whole Seed and Tissue-Specific Food Reserves Reveals a Close Link between the Abundance of Seed Reserves and Seed and Seedling Biomass (in preparation).

- **Khan N.**, Kazmi R.H., Willems L.A.J., Ligterink W., Hilhorst H.W.M. (2013) Canonical Association Reveals a Strong Link between Metabolic Signatures of Seed and Seedling Quality in a Recombinant Inbred Population of Tomato (in preparation)

- **Khan N.**, Willems L.A.J., Ligterink W., Hilhorst H.W.M. (2013) Using Heterogeneous Inbred Families (HIFs) to Confirm Natural Allelic Variation for Complex Seed and Seedling Phenotypes on Tomato Chromosomes 6 and 9 (in preparation)

Other Publications

- Ghulam, R., **Khan, N**., Hashim, M. Malik, Naveed, K. (2007) Comparative Study of Some Promising Sugarcane Varieties at ARI, D.I. Khan. Pak. Sugar J., XXII (05): 23-25

- Mujtaba, M., Shamsi, I.H., **Khan, N.** (2003) Impact of Row spacing and Fertilizer Levels (Diammonium Phosphste) on the Yield and Yield Components of Canola (Brassica napus L). Asian J.of P. Sci. 6(2):454-456.2003

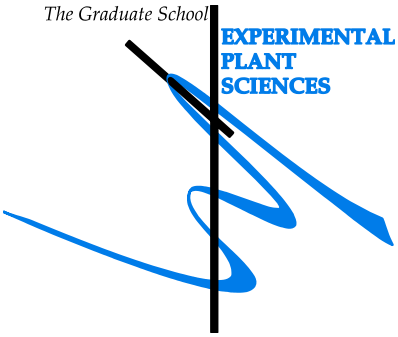
- **Khan N.**, Jan A., Khan I., Khan, I.J., Naeem K. (2002) Response of Canola (Brassica napus L) to Nitrogen and Sulphur Nutrition. Asian J.of P. Sci. 5(1):516-518.2002

- Amanuulah J., **Khan, N.**, Naeem K., Khan I.A., Khattak, B. (2002) Chemical Composition of Canola (Brassica napus L) as affected by Nitrogen and Sulphur. Asian J.of P. Sci. 5(1):519-521.2002.

- Ihsanullah K., Taj F.H., Akbar H., Basir A., **Khan N.** (2002) Effect of Row Spacing on Agronomic Traits and yield of Mungbean (Vigna radiata L.Wileczek). Asian J.of P. Sci. 4(1):328-329.2002.

Education Statement of the Graduate School

Experimental Plant Sciences



Issued to: Noorullah Khan
Date: 3 September 2013
Group: Plant Physiology, Wageningen University

1) Start-up phase	*date*
► **First presentation of your project**	
Genes for seed quality: Genetic and Physiological Quality of Tomato Seed and Seedlings	Mar 30, 2009
► **Writing or rewriting a project proposal**	
Genetic and Physiological Quality of Tomato Seed and Seedlings	Feb 19, 2009
► **Writing a review or book chapter**	
► **MSc courses**	
Genomics: course code ABG - 30306	Feb-Mar, 2009
► **Laboratory use of isotopes**	

Subtotal Start-up Phase *13.5 credits**

2) Scientific Exposure	*date*
► **EPS PhD Student Days**	
EPS PhD student day, Leiden University	Feb 26, 2009
EPS PhD student day, Utrecht University	Jul 01, 2010
EPS PhD student day, Wageningen University	May 20, 2011
EPS PhD student day, University of Amsterdam	Nov 30, 2012
► **EPS Theme Symposia**	
EPS Theme 1 Symposium 'Deveolpmental Biology of Plants', Leiden University	Jan 30, 2009
EPS Theme 1 Symposium 'Developmental Biology of Plants', Wageningen University	Jan 28, 2010
EPS Theme 3 Symposium 'Metabolism and Adaptation', Leiden University	Feb 19, 2010
EPS Theme 3 Symposium 'Metabolism and Adaptation', Wageningen University	Feb 10, 2011
EPS Theme 1 Symposium 'Developmental Biology of Plants', Wageningen University	Jan 19, 2012
EPS Theme 3 Symposium 'Metabolism and Adaptation', Utrecht University	Apr 26, 2012
► **NWO Lunteren Days and Other National Platforms**	

ALW meeting 'Experimental Plant Sciences', Lunteren	Apr 06-07, 2009
ALW meeting 'Experimental Plant Sciences', Lunteren	Apr 19-20, 2010
ALW meeting 'Experimental Plant Sciences', Lunteren	Apr 04-05, 2011
ALW meeting 'Experimental Plant Sciences', Lunteren	Apr 02-03, 2012
► **Seminars (Series), Workshops and Symposia**	
Invited Seminars (Jian-Kang Zhu, Sjef Smeekens)	Nov, 2008
Symposium: The Schilperoort Lectures: Success Stories of Entrepreneurial Scientists, Wageningen University	Nov 05, 2008
Seminar: Science From an Editor's View, by Dr. Pamela J. Hines, Wageningen University	Nov 06, 2008
Symposium: New Opportunities for Conservation Genetics with Genome Wide Information, Wageningen University	Dec 08, 2008
Invited Seminar Hiro Nonogaki	Sep 17, 2009
EPS symposium 'Ecology and Experimental Plant Sciences 2', Wageningen University	Sep 22, 2009
Farewell Symposium Pim Zabel ''Art Meets Science'', Wageningen University	Dec 07, 2011
Invited Seminars (John Yoder, David Baulcombe)	Aug-Sep, 2010
Seminar SNIP Detection by Allumina Sequencing, Robert Kraus	Oct 07, 2010
Wageningen UR Sequencing Seminar, Wageningen University	Dec 07, 2011
Invited Seminars (Steffen Abel, Graham Seymour)	Mar-Sep, 2011
Invited Seminars (Steven Penfield, Jill M. Farant, Lauren McIntyre, Aaron Fait)	Jan-Dec, 2012
Seminar series Plant Physiology-Genetics, Wageningen University	2008-2012
Workshop 'Transcriptome'	Jun 26, 2013
► **Seminar Plus**	
► **International Symposia and Congresses**	
3rd Workshop on Molecular Aspects of Seed Dormancy and Germination, York, UK	Jul 18-21, 2010
2nd Int. Joint PhD Retreat, Cologne, Germany	Apr 15-17, 2010
10th ISSS Workshop on Seed Biology, Salvador (Bahia State), Brazil	Apr 13-18, 2011
3rd Int. Joint PhD Retreat, Orsay, France	Jul 05-08, 2011
► **Presentations**	
Poster: 3rd Joint PhD Retreat, Orsay, France	Apr 15-17, 2010
Poster: ALW meeting 'Experimental Plant Sciences', Lunteren	Apr 19-20, 2010
Poster: ALW meeting 'Experimental Plant Sciences', Lunteren	Apr 04-05, 2011
Poster: ALW meeting 'Experimental Plant Sciences', Lunteren	Apr 02-03, 2012
Oral: 3rd Workshop on Molecular Aspects of Seed Dormancy and Germination, York, UK	Jul 18-21, 2010
Oral: 10th ISSS Workshop on Seed Biology	Apr 13-18, 2011
Oral: 3rd Int. Joint PhD Retreat, Orsay, France	Jul 05-08, 2011

► IAB Interview	Feb 19, 2011
► Excursions	
Subtotal Scientific Exposure	*21.6 credits**

3) In-Depth Studies	***date***
► EPS courses or other PhD courses	
Basic statistics course, Wageningen University	Dec 15-22, 2009
6th International Master Class on Seed Technology, Wageningen University	Oct 26-29, 2009
► Journal club	
Literature discussion: Plant physiology group, Wageningen University	2008-2012
► Individual research training	
Subtotal In-Depth Studies	*5.7 credits**

4) Personal development	***date***
► Skill training courses	
PhD Competence Assessment, Wageningen University	Apr, 2009
PhD Course 'Information Literacy for PhD, including introduction EndNote', Wageningen University	Jun 09-10, 2009
PE&RC Day 'Selling Science', Wageningen University	Oct 28, 2010
EPS Career Event: ExPectationS Day, Wageningen University	Nov 19, 2010
Techniques for Writing and Presenting a Scientific Paper	Jul 02-05, 2013
► Organisation of PhD students day, course or conference	
► Membership of Board, Committee or PhD council	
Subtotal Personal Development	*2.7 credits**

TOTAL NUMBER OF CREDIT POINTS*	43.5

Herewith the Graduate School declares that the PhD candidate has complied with the educational requirements set by the Educational Committee of EPS which comprises of a minimum total of 30 ECTS credits

** A credit represents a normative study load of 28 hours of study.*

Printed by: Proefschriftmaken.nl || Uitgeverij BOXPress