



Plasticity of maternal environment-dependent expression-QTLs of tomato seeds

Mark G. Sterken¹ · Harm Nijveen^{2,3} · Martijn van Zanten⁴ · Jose M. Jiménez-Gómez^{5,6} · Nafiseh Geshnizjani² · Leo A. J. Willems² · Juriaan Rienstra² · Henk W. M. Hilhorst² · Wilco Ligterink² · Basten L. Snoek^{1,7}

Received: 27 December 2021 / Accepted: 27 October 2022 / Published online: 22 February 2023
© The Author(s) 2023

Abstract

Seeds are essential for plant reproduction, survival, and dispersal. Germination ability and successful establishment of young seedlings strongly depend on seed quality and on environmental factors such as nutrient availability. In tomato (*Solanum lycopersicum*) and many other species, seed quality and seedling establishment characteristics are determined by genetic variation, as well as the maternal environment in which the seeds develop and mature. The genetic contribution to variation in seed and seedling quality traits and environmental responsiveness can be estimated at transcriptome level in the dry seed by mapping genomic loci that affect gene expression (expression QTLs) in contrasting maternal environments. In this study, we applied RNA-sequencing to construct a linkage map and measure gene expression of seeds of a tomato recombinant inbred line (RIL) population derived from a cross between *S. lycopersicum* (cv. MoneyMaker) and *S. pimpinellifolium* (G1.1554). The seeds matured on plants cultivated under different nutritional environments, i.e., on high phosphorus or low nitrogen. The obtained single-nucleotide polymorphisms (SNPs) were subsequently used to construct a genetic map. We show how the genetic landscape of plasticity in gene regulation in dry seeds is affected by the maternal nutrient environment. The combined information on natural genetic variation mediating (variation in) responsiveness to the environment may contribute to knowledge-based breeding programs aiming to develop crop cultivars that are resilient to stressful environments.

Mark G. Sterken and Basten L. Snoek have equally contributed to this work.

✉ Wilco Ligterink
wilcoligterink@gmail.com

✉ Basten L. Snoek
l.b.snoek@uu.nl

Mark G. Sterken
mark.sterken@wur.nl

Harm Nijveen
harm.nijveen@wur.nl

Martijn van Zanten
m.vanzanten@uu.nl

Jose M. Jiménez-Gómez
Jose.Jimenez-Gomez@inrae.fr

Nafiseh Geshnizjani
nafiseh.geshnizjani@gmail.com

Leo A. J. Willems
leo.willems@wur.nl

Juriaan Rienstra
juriaan.rienstra@wur.nl

Henk W. M. Hilhorst
henk.hilhorst@wur.nl

- 1 Laboratory of Nematology, Wageningen University, 6708 PB Wageningen, The Netherlands
- 2 Wageningen Seed Lab, Laboratory of Plant Physiology, Wageningen University, 6708 PB Wageningen, The Netherlands
- 3 Laboratory of Bioinformatics, Wageningen University, 6708 PB Wageningen, The Netherlands
- 4 Plant Stress Resilience, Institute of Environmental Biology, Utrecht University, 3584 CH Utrecht, The Netherlands
- 5 Department of Plant Breeding and Genetics, Max Planck Institute for Plant Breeding Research, Cologne, Germany
- 6 Institut Jean-Pierre Bourgin, INRAE, AgroParisTech, Université Paris-Saclay, 78000 Versailles, France
- 7 Theoretical Biology and Bioinformatics, Institute of Biodynamics and Biocomplexity, Utrecht University, 3584 CH Utrecht, The Netherlands

Introduction

Seeds are essential for reproduction and dispersal of plants and function as survival structures to overcome harsh environmental conditions unfavorable for plant growth. Well-timed development and ripening of seeds, to ensure optimal seed performance and the ability to germinate in a permissive environment, are therefore essential for plant fitness. Successful germination strongly depends on seed performance, which is affected by environmental conditions, such as temperature, water availability, light conditions, and the nutrient status that the maternal plant experienced (de Souza et al. 2020; Delouche and Baskin 1971; Delouche 1980; Donohue 2009; Rowse and Finch-Savage 2003). More specifically, seed performance/germination in species such as tomato and the model plant *Arabidopsis thaliana* is determined during seed development and maturation and depends on temperature (Demir et al. 2004; He et al. 2014; Schmuths et al. 2006), photoperiod (Munir et al. 2001; Pourrat and Jacques 1975), nutrient composition, and levels (Alboresi et al. 2005; Geshnizjani et al. 2019; He et al. 2014). Seed quality, germination, and seedling establishment traits also have strong genetic determinants, and (natural) genetic variation in quality traits, including Quantitative Trait Loci (QTLs), has been reported (Geshnizjani et al. 2019, 2020; He et al. 2014; Joosen et al. 2012; Khan et al. 2012; Serin et al. 2017).

Phosphate and nitrate are essential plant nutrients with profound effects on plant growth (Schachtman et al. 1998; Urbanczyk-Wochniak and Fernie 2005) and seed performance/germination traits (Alboresi et al. 2005; Geshnizjani et al. 2019, 2020; He et al. 2014). In *Arabidopsis*, it has been shown that seeds produced by plants fertilized with higher-than-normal levels of phosphate showed increased germination rates under stressful conditions (He et al. 2014). Nitrate is known to have a strong effect on seed germination and seed dormancy in multiple plant species (Duermeyer et al. 2018), with high concentrations of nitrate supplied to the mother plant leading to lower dormancy of the seeds (Alboresi et al. 2005). This is attributed to nitrogen effects on the gibberellin/abscisic acid (GA/ABA) balance in the seeds, with higher endogenous nitrate levels resulting in lower ABA levels in seeds and hence shallower dormancy (Matakiadis et al. 2009). In *Arabidopsis*, altered nitrate levels experienced by the mother plant also have a substantial effect on the levels of multiple metabolites and transcripts in the seeds, with a notable reduction in nitrogen metabolism-related metabolites and genes (He et al. 2016).

Tomato (*Solanum lycopersicum*) is one of the most important vegetable crops worldwide and is a model

organism for research on fruit-bearing crops (Giovannoni 2001; Schauer et al. 2006; Tomato Genome 2012; Tomato Genome Sequencing et al. 2014). However, in the process of domestication, breeding selection and propagation, a substantial fraction of the genetic variation in the founder's germplasms has been lost (Razifard et al. 2020; Tomato Genome 2012; Tomato Genome Sequencing et al. 2014). Moreover, due to a focus on fruit quality, resistance, and yield traits, other desirable traits that have not been directly selected for could have been lost over time in modern varieties (McCouch 2004; Razifard et al. 2020; Wang et al. 2020). In plants other than tomato, this includes several seed quality traits (Bauchet et al. 2014; Doebley et al. 2006; McCouch 2004; Razifard et al. 2020; Wang et al. 2020). Trait variation loss could be restored by including wild cultivars/ancestors of modern commercial tomato such as *Solanum pimpinellifolium*. Yet, although wild ancestors represent a rich source of genetic variation (although also underlying negative traits) in breeding programs and in studies on tomato (quantitative) genetics (Blanca et al. 2015; Lin et al. 2014; Pascual et al. 2016; Razali et al. 2018; Tomato Genome 2012; Yang et al. 2014). For instance, wild cultivars have been used in genetic screens and genome wide association studies (GWAS) to discover genomic loci and genes involved in variation in metabolic traits (Bauchet et al. 2017; Sauvage et al. 2014; Ye et al. 2019; Zhang et al. 2015; Zhao et al. 2019), insect resistance (Vosman et al. 2018), floral meristem identity (Bauchet et al. 2014), trichome formation (Chang et al. 2018), microbial rhizosphere composition (Oyserman et al. 2022), and fruit shape and size (Albert et al. 2016; Blanca et al. 2015; Razifard et al. 2020). In addition to GWAS, Recombinant Inbred Line (RIL) populations, derived from experimental crossing between *S. lycopersicum* and *S. pimpinellifolium*, are frequently used to uncover the effect of genetic variation on tomato traits (Capel et al. 2015, 2017; Celik et al. 2017; Kazmi et al. 2017; Viquez-Zamora et al. 2014; Voorrips et al. 2000; Zhang et al. 2018), including various seed quality traits (de Souza et al. 2016; Geshnizjani et al. 2018, 2019, 2020; Khan et al. 2012).

The introduction and improved feasibility of diverse omics techniques have accelerated studies into the molecular mechanisms underlying natural variation in tomato traits in the past two decades (Rothan et al. 2019). In particular, advances in transcriptomics techniques such as microarray analysis and later RNA-sequencing have proved useful in this context, by enabling, e.g., GWAS studies. Moreover, measuring gene expression in RILs has enabled expression-QTL (eQTL) analysis as a powerful tool to detect gene regulatory loci (Jansen and Nap 2001; Jimenez-Gomez et al. 2010; Kawakatsu et al. 2016; Keurentjes et al. 2007; Snoek et al. 2012; West et al. 2007). Combining the wealth

of information obtained by mapping eQTLs enables (re) construction of regulatory networks underlying plant traits (Jimenez-Gomez et al. 2010; Keurentjes et al. 2007; Terpstra et al. 2010). In addition, comparison of eQTL profiles from multiple environments may aid our understanding of how genetic variation shapes the effects the environment has on the appearance of phenotypes (Hartanto et al. 2020; Nijveen et al. 2017; Snoek et al. 2012). In plant (*Arabidopsis*) and worm (*Caenorhabditis elegans*) model systems, it has been shown that especially *trans*-eQTLs are dynamic and can be highly specific for a certain environment (Cubillos et al. 2014; Hartanto et al. 2020; Nijveen et al. 2017; Snoek et al. 2017, 2012; Sterken et al. 2019; Vinuela et al. 2010).

Although seed quality and seedling establishment characteristics are determined by both genetic variation and the maternal environment in which the seeds develop and mature (Geshnizjani et al. 2019, 2020; He et al. 2014), it is currently unknown if the maternal environment causes a perturbed eQTL landscape in the progeny seeds and how the nutrient environment of the mother plant affects these landscapes. We therefore followed an RNA-seq approach and quantified natural variation in mRNA levels in the dry seeds of a tomato RIL population from a cross derived from *S. lycopersicum* (cv. MoneyMaker) and *S. pimpinellifolium* (G1.1554) parents (Khan et al. 2012; Voorrips et al. 2000) that were cultivated either in a low nitrogen or a high phosphorus environment. In this work, we first present a high-density RNA-seq-derived genetic map of tomato, and subsequently, we demonstrate how the genetic landscape of gene regulation of tomato dry seeds is affected by the nutritional environment of the mother plant. Altogether, our detailed analysis of the genetic underpinning of plasticity in gene expression as responsiveness to the maternal environment, attributed to the progeny seeds, may contribute to knowledge-based breeding programs aiming to develop crop cultivars that are resilient to stressful environments, including production of high-quality seeds under sub-optimal environmental conditions.

Results

An RNA-seq-derived genetic map of tomato

We performed an RNA-sequencing experiment to uncover the interplay between genetic variation, the nutritional status of the maternal environment, and mRNA abundances in progeny tomato seeds. The used seeds were derived from 101 tomato RIL plants of a cross between *S. lycopersicum* (cv. MoneyMaker; MM) and *S. pimpinellifolium* (G1.1554 or CGN14498; PI) (Kazmi et al. 2012; Voorrips et al. 2000) and their parental lines. All maternal plants were pre-cultivated on standard nutrient conditions and upon

flowering transferred to either low nitrogen (LN; 52 RILs) or high phosphate (HP; 49 RILs) nutrition. The two RIL sets were non-overlapping (Geshnizjani et al. 2020).

In addition to estimating expression differences among individuals, RNA-seq reads allowed for the identification of single-nucleotide polymorphisms (SNPs) in transcribed genes of the parental lines and the RILs. These SNPs were subsequently used to construct genetic and physical maps of the RIL population, to facilitate QTL and eQTL mappings (Serin et al. 2017; Snoek et al. 2019). In total, we detected 43,188 consistent SNPs between the parental lines. These SNPs were subsequently used to reconstruct the genotypes (*i.e.*, determine the crossover locations) of the RILs (Fig. 1a, b). Across our RIL set, a balanced distribution of the parental alleles was observed genome-wide, with the notable exception of chromosome 2, which had a substantial higher frequency of PI alleles (Supplementary Fig. 1). Measured over all RILs, 2847 recombination (crossover) events were detected. As expected, the crossovers were found almost exclusively in euchromatic regions of the chromosomes, causing severe distortion between the physical and genetic maps, as described before (Demirci et al. 2017) (Fig. 1c). On average, two recombination events were detected per RIL per chromosome. Altogether, the population size and recombination events provided 4515 unique genetic markers and 4568 distinguishable genomic loci/bins suitable for mapping, improving the previously available map (Kazmi et al. 2012) (Supplementary Table 1). The detected loci had a size range from 60 to 1.7 Kb, with an average locus size of 180 Kb and a median of 11 Kb (Supplementary Table 2). Given the high local recombination frequency, relatively small loci were overrepresented toward the chromosome tips (Fig. 1c). Together, our dataset enables precise mapping of QTLs and eQTLs, especially toward the tips of the chromosomes but with poor resolution near the centromeres.

The maternal nutrient environment affects mRNA abundances in seeds

Next, we compared mRNA abundances in all HP-treated lines (RILs and parental lines) with the mRNA abundances in LN-treated lines, to identify genes contributing to differences between the two environments. Principal Component Analysis (PCA) demonstrated the presence of a substantial effect of the maternal nutrient environment on transcript levels in seeds (Fig. 2a). A linear model was used to identify which mRNAs were differentially expressed between the two maternal environments. A multiple testing correction was applied, and differential expression of 2871 mRNAs (out of 14,772 detected mRNAs) was found (Bonferroni corrected p -value < 0.05) to depend on the nutritional conditions the mother plant experienced during the seed maturation phase (*i.e.*, LN or HP) (Supplementary Table 3). Of these 2871

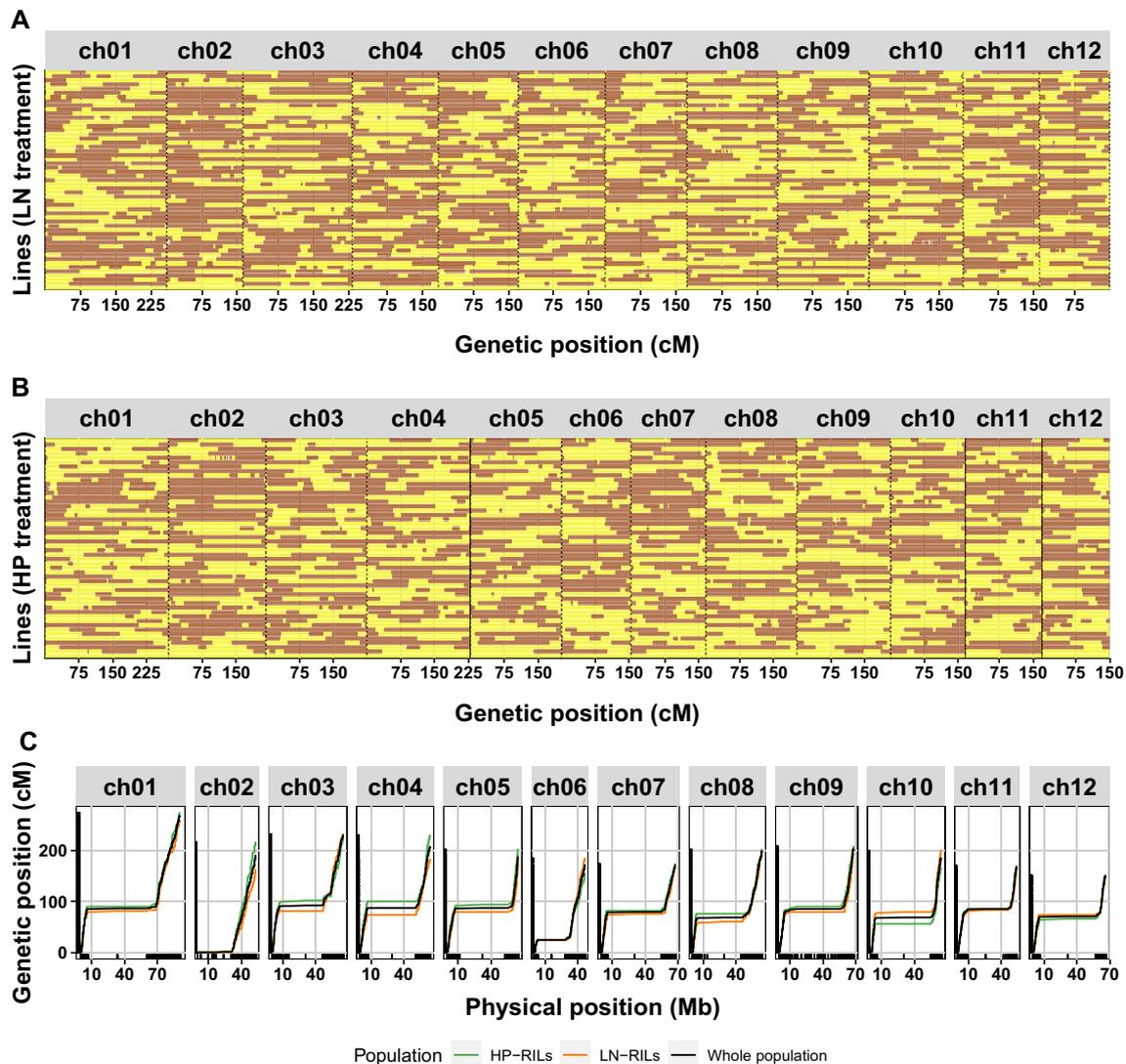


Fig. 1 **a** Genetic map showing the genotypes of the RILs that were included in the low nitrogen (LN) treatment experiment. The map represents the most probable genotype per marker location. Yellow indicates MM; brown indicates PI. Position on the chromosome is indicated in centimorgans (cM). **b** As in (a), but for the RILs that were included in the high phosphorous (HP) treatment experiment.

c Recombination events per chromosome for the whole population (black) and the LN-(orange) and HP-treated (green) sub-populations. The physical position is given on the x-axis in million bases (Mb); the genetic position is given on the y-axis in centimorgans (cM). Chromosome numbers are indicated above panels (color figure online)

mRNAs, 922 were more abundant in seeds developed and ripened in HP conditions compared to LN, and 1949 mRNAs were significantly more abundant in LN conditions compared to HP. The mRNAs of genes that were more abundant after LN treatment were among others enriched for Gene Ontology (GO) terms: 'chloroplast,' 'ATP binding,' 'proteasome,' and 'nitrate transport' (Supplementary table 4). mRNAs that were more abundant in seeds grown in HP conditions were enriched for the GO terms: 'cellular response to hypoxia,' 'pectin esterase activity,' and 'glucosinolate metabolic process' (Supplementary Table 4).

We also inquired the differences of the mRNA abundances between the MM and PI parental lines, within and between treatments. To this end, we again employed a linear model, but were less stringent in the statistical thresholds (as there were no confounding effects). We found 2976 mRNAs differentially expressed between the two parental lines regardless of treatment and 382 mRNAs that were differentially expressed between the lines due to treatment (linear model, $FDR \leq 0.05$; Fig. 2b and Supplementary table 5). GO enrichment indicated that the 1240 mRNAs more abundant in MM compared to PI were, among other categories, enriched, for 'transcription factor activity,'

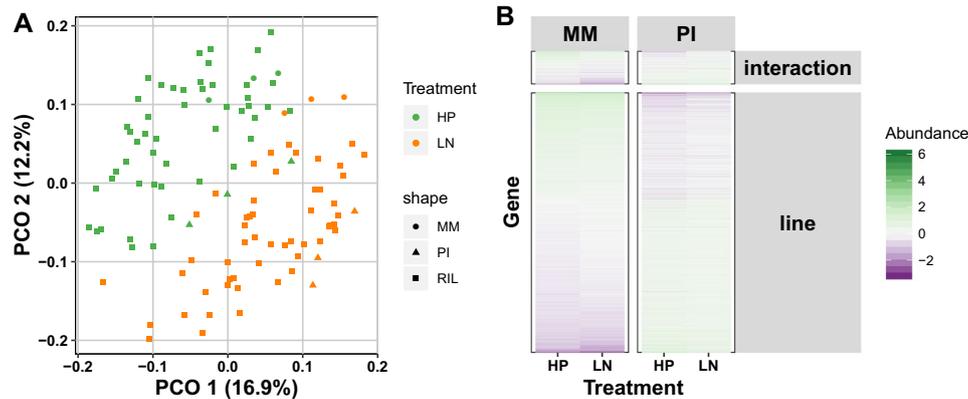


Fig. 2 Nutrition status-related mRNA abundance differences. **a** The first two axes of a principal component analysis on the \log_2 ratio with the mean transcripts per million (TPM) values. The first axis (PCO1) explained 16.9% of the variance in the data, the second 12.2%. Square symbols represent individual RILs; Moneymaker (MM) parental samples are represented by dots and *S. pimpinellifolium* (PI) paren-

tal samples by triangles. The colors indicate high phosphorous (HP; green) or low nitrogen (LN; orange) treatments. **b** Differentially abundant mRNAs in the two parental lines not affected by treatment (2976 genes, line) and affected by treatment (382 genes, interaction). Lower abundance is shown in purple and higher in green. The genes can be found in Supplementary Table 5 (color figure online)

‘oxidation–reduction,’ ‘protein -binding,’ ‘-phosphorylation,’ ‘-ubiquitination,’ ‘chloroplast,’ ‘circadian rhythm,’ and ‘metal ion binding’ (Supplementary Table 6a). The 1736 mRNAs that were more abundant in PI compared to MM were, among other categories, enriched for ‘cytosol,’ ‘chloroplast,’ ‘nucleus,’ ‘mitochondrion,’ ‘cytoplasm,’ ‘ribosome,’ ‘translation,’ ‘nucleolus,’ ‘endoplasmic reticulum,’ ‘oxidation–reduction,’ ‘vacuole,’ and ‘copper ion binding’ (Supplementary table 6a). The 382 genes showing a significant interaction effect between the parental background and maternal environment showed an enrichment for the GO terms ‘oxidation–reduction,’ ‘extracellular region,’ ‘transcript regulation,’ ‘iron ion binding,’ and ‘response to gibberellin’ (Supplementary table 6b). Of note, the ‘oxidation–reduction process’ and ‘transcript regulation’ GO terms are enriched in the upregulated genes of both MM and PI, which is not surprising since both GO terms are quite general and each represents many genes. These results show that the nutrition status of the mother plant (environment; E) as well as genotype (G), and the interaction between the two ($G \times E$), modulates mRNA abundances in dry seeds of tomato.

Heritability and transgression in mRNA abundances

To estimate the contribution of genetic variation to differences in mRNA abundance between the genetic backgrounds (plant lines) and treatments (nutrient status), we calculated the Broad-Sense Heritability (BSH). It should be noted that the method used in general gives an upper-bound estimation of the BSH (Brem and Kruglyak 2005; Keurentjes et al. 2007; Rockman et al. 2010; Snoek et al. 2012; Sterken et al. 2019). In addition, replicated

tal samples by triangles. The colors indicate high phosphorous (HP; green) or low nitrogen (LN; orange) treatments. **b** Differentially abundant mRNAs in the two parental lines not affected by treatment (2976 genes, line) and affected by treatment (382 genes, interaction). Lower abundance is shown in purple and higher in green. The genes can be found in Supplementary Table 5 (color figure online)

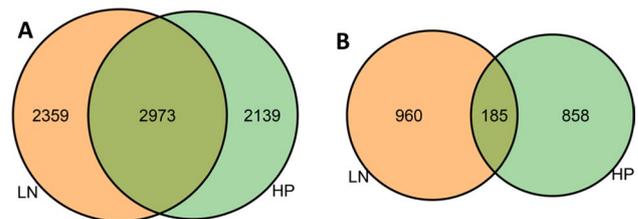


Fig. 3 Venn diagrams showing the overlap and differences of **a** genes with significant heritability and **b** genes exhibiting significant transgression, of mRNA abundance levels between LN (orange) and HP (green; $FDR < 0.05$) (color figure online)

measurements in the parental lines were used to estimate non-genetic variance. We found 5112 genes in HP and 5332 genes in LN that showed significant heritability for mRNA abundance, of which 2973 genes overlapped (39.8%; permutation, $FDR < 0.05$; Fig. 3a; Supplementary Table 7a). Subsequently, we checked if genes with significant heritable contribution to mRNA abundance differences were predominantly affected by the maternal nutrient environment. However, we did not find such an enrichment for any of the overlapping groups of genes (hypergeometric test, $p > 0.01$; Supplementary Fig. 2a). We thus conclude that, overall, the number of genes with significant heritability for mRNA abundance was not specifically responsive to the maternal nutrient treatments. The genes with heritable mRNA abundance in HP alone were enriched for the GO terms: ‘translation,’ ‘ribosome,’ ‘mitochondrion,’ and more (Supplementary Table 7b). Those that showed significant heritability only in LN were enriched for the GO terms: ‘ABA metabolic process,’ and others (Supplementary table 7b). The genes

that showed significant heritability in both environments were enriched for various GO terms: ‘oxidation–reduction process,’ ‘ribosome/translation,’ ‘nucleolus,’ ‘cell wall,’ ‘heme binding,’ ‘ion binding,’ and ‘vacuole’ (Supplementary Table 7b).

We next assessed the complexity of the genetic regulation underlying mRNA abundance differences. To this end, the transgression was calculated, i.e., trait values in RILs that extent beyond the parental means. We found significant transgression in mRNA abundance (trait) levels for 1043 genes in the maternal HP treatment and 1145 genes in the maternal LN treatment (permutation, $FDR < 0.05$; Supplementary Table 8a). This suggests a polygenic genetic architecture for mRNA abundance. Of these, the mRNA abundances of 185 genes showed significant transgression beyond the parental means in both treatments (Fig. 3b). Also, here, we tested for significant overlap with treatment-related genes. Yet, with 18% response to treatment of the transgressive mRNAs, there was no significant enrichment for transgressive mRNA abundances with treatment-related differences (hypergeometric test, $p > 0.01$; Supplementary Fig. 2b). So, alike heritability, transgression is apparently not linked to a reduction of nitrogen or increase in phosphorus content in the maternal growth environment. Moreover, compared to genes showing significant heritability, many fewer GO terms were enriched in the genes showing transgression, and those GO terms that were enriched generally had a lower level of significance. For genes showing transgression in HP alone, the GO terms ‘cell periphery,’ ‘positive gravitropism,’ ‘cysteine biosynthetic process,’ ‘symporter activity,’ and ‘response to heat’ were enriched, whereas for genes only showing transgression in LN the GO-terms ‘beta-glucosidase activity,’ ‘preprophase band,’ and ‘phragmoplast’ were enriched. The GO term ‘DNA-binding transcription factor activity’ was enriched in genes showing transgression in both environments (Supplementary Table 8b).

The maternal nutrient environment produces specific eQTL landscapes

Altogether, our analyses revealed both a considerable effect of the maternal nutrient environment (HP versus LN) and a significant influence of genetic variation in the RIL panel (heritability) on the detected mRNA abundance levels. By combining our constructed SNP genetic map (Fig. 1a, Supplementary table 1) with the obtained mRNA abundance dataset (Fig. 2), we were able to identify eQTLs that potentially contribute to the variation in mRNA abundance (Fig. 4a–f). In other words, the identified eQTL have a high chance of harboring polymorphic regulatory factors (e.g., genes or other genetic elements) for mRNA abundance,

prospectively explaining variation in the seed and germination trait phenotypes observed.

We detected a maternal environment-specific *trans*-eQTL landscape, as the distribution of the position of the *trans*-eQTLs was very different between the two environments. For the HP environment, 4281 eQTLs for 3833 genes were identified, of which 2247 were *cis*-eQTLs and 2034 were *trans*-eQTLs. For the LN environment, 7487 eQTLs were detected for 6815 genes, of which 2356 were *cis*-eQTLs and 5131 were *trans*-eQTLs ($FDR < 0.05$; $-\log_{10}(p) > 3.9$; Fig. 4a–d; Supplementary Table 9; Supplementary Table 10). The confidence intervals of the eQTL locations were mostly dependent on the chromosomal area where the QTL was mapped, with small intervals at the chromosomal arms and large intervals nearer the centromeres (Supplementary Table 9). A significant overlap between *cis*-eQTLs of the two environments was noted (Fig. 4e; 1506 overlapping *cis*-eQTLs; 48.6%; hypergeometric test, $p < 1 \cdot 10^{-16}$). On the contrary, the *trans*-eQTLs were mainly specific for each tested maternal environment (Fig. 4f; 590 overlapping *trans*-eQTLs; 9.7%; hypergeometric test, $p = 1.0$). However, both *cis*- and *trans*-eQTLs were not enriched for genes with differentially abundant mRNA levels based on the maternal environment (hypergeometric test, $p > 0.01$; Supplementary Fig. 3a and b). Together with the significant transgression (Supplementary Table 8a) and considerable heritability of mRNA abundances (Supplementary Table 7a; Fig. 3a), this indicates that *trans*-eQTLs represent a genotype-specific interaction with the maternal nutrient environment. Many different GO terms were found to be enriched in the genes with environment-specific eQTLs. For an overview, see Supplementary table 11.

The majority of the *trans*-eQTLs clustered in maternal nutrient environment-specific eQTL hotspots or *trans*-bands (Fig. 4a, c). Hence, these genomic regions harbor the main loci underlying the genetic variation in environment-specific gene expression regulation in our dataset. A total of 13 *trans*-bands (9 in the HP treatment and 4 in the LN treatment; see Methods for the *trans*-band criteria) were identified, which account for 1206 of the *trans*-eQTLs in the HP treatment (59.3% of HP total) and 4181 of the *trans*-eQTLs in the LN treatment (81.5% of LN total; Table 1).

Thus, *trans*-bands are a major explanatory factor for *trans*-eQTLs. In other words, a relatively large proportion of *trans*-eQTLs are caused by a few pleiotropic major effect loci. Remarkably, the MM allele had a positive effect on mRNA abundance for the majority of the eQTLs of the *trans*-bands in the LN soil environment, whereas this was not so prevalent in the HP environment (Table 1; Fig. 4a,

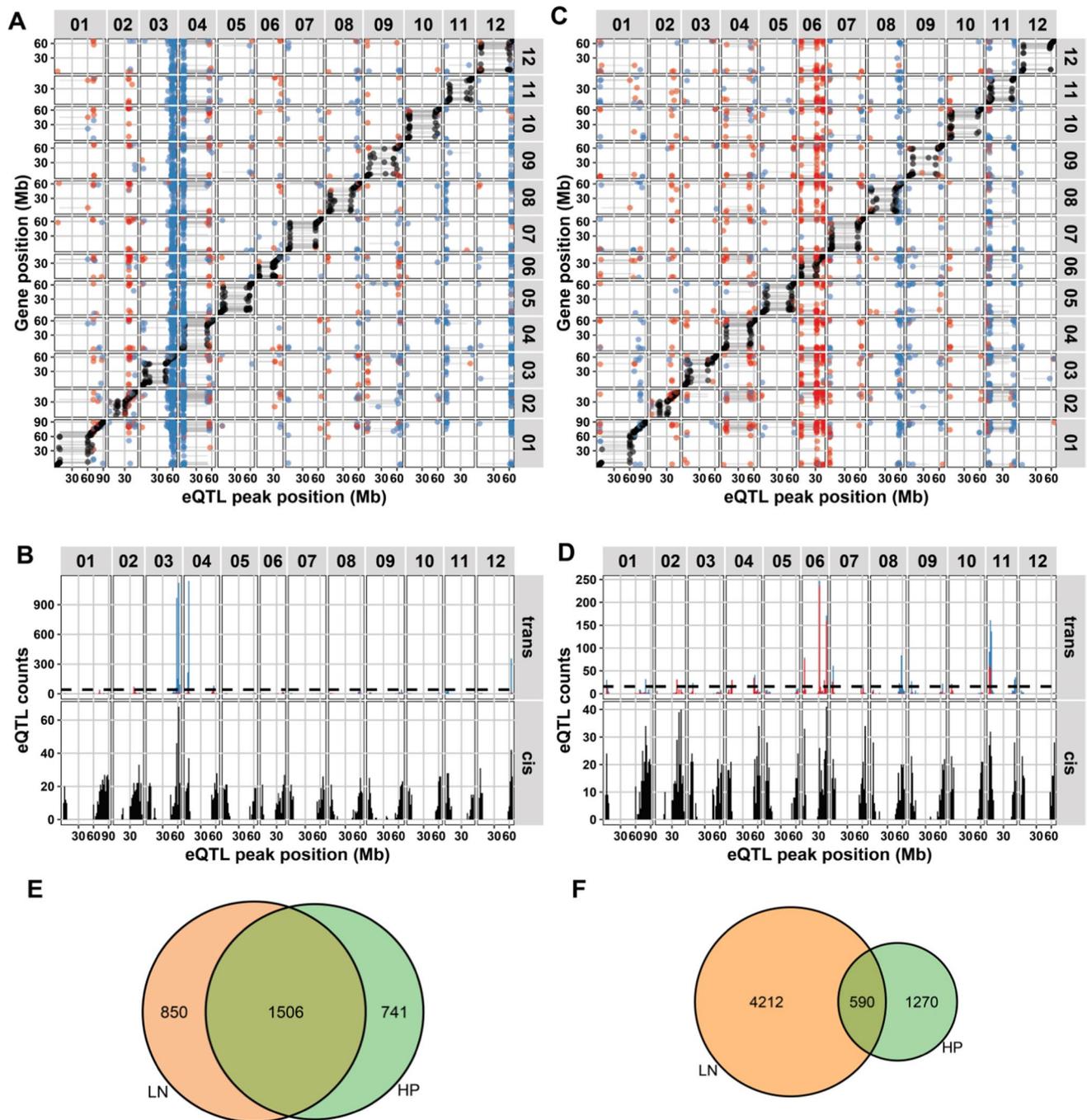


Fig. 4 Characteristics of the detected eQTL landscapes in tomato dry seeds in **a, b** LN and **c, d** HP environments. **a, c** *Cis-trans* plots of eQTLs mapped ($-\log_{10}(p) > 3.9$). The positions of the eQTL peaks are plotted on the *x*-axis and the positions of the corresponding genes on the *y*-axis. Chromosomes are indicated on the top and right in the gray labels. Colors indicate *cis*-eQTL (black), eQTL associated with higher mRNA abundance due to the MM allele (blue) or with higher

c. Most of these *trans*-bands showed enrichment for specific GO terms, such as ‘translation’ and ‘specific cellular

abundance by the PI allele (red). **b, d** Histograms showing the distribution of the *cis*- and *trans*-eQTL over the chromosomes, arranged by eQTL peak location counted per 2 million bases (Mb) bins. The dashed lines in the *trans*-eQTL panels indicate the threshold for calling a *trans*-band (Poisson distribution, $p < 0.0001$). **e** The overlap of *cis*-eQTL in the two treatments and **f** the overlap of *trans*-eQTL in the two maternal environments (color figure online)

organelles’ for LN and ‘oxidoreductase’ and ‘vacuole’ for HP (Table 1, Supplementary Table 12). Moreover, many of the *trans*-bands co-locate with known QTLs for germination and seed traits [Table 1 (Geshnijani et al. 2020; Khan et al.

Table 1 Overview of detected *trans*-band (TB) eQTLs. Indicated are given ID's, location on the physical genome (map position in Mb), number of eQTLs underlying the *trans*-band (+sign: MM>PI; –sign PI>MM), GO terms enriched in the eQTLs underlying the *trans*-band in either MM or PI and co-location with known phenotypic QTLs for germination (Geshnizjani et al. 2020; Khan et al. 2012); +sign: MM>PI; –sign PI>MM)

TB ID	Position	eQTLs	GO enrichment (+MM higher)	GO enrichment (–PI higher)	Germination QTL Refs: (Geshnizjani et al. 2020)
LN_TB1	ch03: 56–64 Mb	2369 (2330+; 39–)	Translation; ribosome; nucleolus; RNA binding; mitochondrion; cell wall; and more	None	Th-I [57.5 Mb] (LN–)
LN_TB2	ch04:4–8 Mb	1348 (1311+; 37–)	Telomere; nucleus; protein binding; ubiquitin; and more	None	Gmax water [3.14 Mb] (LN+); T10 water [3.14 Mb] (LN+); T10 NaCl [3.14 Mb], mann [4.94 Mb], HT [3.14 Mb] (LN+); T50 water [3.14 Mb], mann [4.94 Mb], (LN+); AUC water [3.14 Mb], mann [3.14 Mb] (LN+)
LN_TB3	ch04:54–56 Mb	99 (45+; 54–)	None	Secretory vesicle; and more	Gmax NaCl [55.0 Mb], mann [58.2 Mb] (LN+); SW [50.2 Mb] (LN+)
LN_TB4	ch12:62–66 Mb	365 (348+; 17–)	Golgi; endosome; glycosylation; ER; and more	None	None in LN
HP_TB1	ch01:2–4 Mb	41 (16+; 25–)	None	None	None in HP
HP_TB2	ch04:6–8 Mb	38 (7+; 31–)	None	Heme binding; oxidoreductase; iron ion binding	Gmax mann [3.14 Mb], HT[3.14 Mb] (HP+); Th-I [1.90 Mb] (HP+)
HP_TB3	ch06:2–4 Mb	96 (5+; 91–)	None	None	U8416 NaCl [9.03 Mb], HT[9.03 Mb] (HP+)
HP_TB4	ch06:32–34 Mb	254 (12+; 242–)	None	RNA processing	Th-T [33.7 Mb] (HP+)
HP_TB5	ch06:44–48 Mb	182 (13+; 169–)	None	Tricarboxylic acid cycle; plastid; vacuolar membrane; cell wall	T10 NaCl [43.8 Mb], mann [43.8 Mb], HT [43.6b] (HP+); T50 NaCl [43.8 Mb], HAT [43.8 Mb] (HP+); AUC NaCl [43.8 Mb] (HP+)
HP_TB6	ch07:0–2 Mb	77 (32+; 45–)	None	None	None in HP
HP_TB7	ch08:58–60 Mb	83 (80+; 3–)	Vacuole; oxidoreductase; golgi	None	Th-D [59.6 Mb] (HP+)
HP_TB8	ch11:0–6 Mb	371 (249+; 122–)	chromosome, centromeric region; ubiquitin conjugating enzyme activity	Transferase activity; hydrolase activity; response to heat	None
HP_TB9	ch11:52–54 Mb	64 (63+; 1–)	Ribosome; nucleolus; translation; and more	None	T10 mann [48.3 Mb] (HP–)

2012)]. These eQTLs can therefore contribute to uncovering the molecular genetic mechanisms underlying the germination and seed trait QTLs.

Discussion

Our RNA-sequencing data obtained from a Tomato RIL population (*S. lycopersicum* (cv. MoneyMaker; MM) × *S. pimpinellifolium* (G1.1554; PI)) (Kazmi et al. 2012; Voorrips et al. 2000) allowed for the construction of a genetic

map, describing the genotypes using 4515 SNP markers. This is over five times more than previously reported in Kazmi et al. (2012), which used 865 markers. However, intrinsic to RNA-seq data, only SNPs present in the coding parts of the genes (mRNA's) could be used. Therefore, determining the exact locus where recombination took place would need additional genome sequencing as described in (Demirci et al. 2017). Furthermore, recombination events were mostly limited to the chromosome arms, leading to a skew in mapping resolution, with more narrow QTLs on

the arms when compared to those located more toward the centromere regions.

By measuring transcript levels (i.e., mRNA abundances) in the seeds of a tomato RIL population that had matured in different maternal nutrient environments, we show that the maternal environment affects both regulation and the genetic architecture of gene expression in progeny seeds. By design of the experiment which focusses on the role of genotypic variation within an environment, we can only observe the differences between the two (extreme) environments. However, this comparison would benefit from the contrasts between normal nutrient conditions and these more extreme environments. An RNAseq experiment on seeds coming from standard nitrogen/phosphorous conditions would surely benefit this analysis and might help identification of gene expression or even eQTL more specific to high nutrient environments.

The genetic analysis revealed that especially *trans* eQTLs proved environment specific, which is comparable to other species (Albert et al. 2018; Cubillos et al. 2014; Hartanto et al. 2020; Li et al. 2006; Nijveen et al. 2017; Snoek et al. 2017, 2020, 2012; Vinuela et al. 2010). We found 3833 genes (~26% of all detected expressed genes in the RILs), with an eQTL in HP and 6815 genes (~46% of all expressed genes in the RILs) with an eQTL in LN. This is comparable to the number detected by Ranjan et al. 2016 (Ranjan et al. 2016), who used the upper part of five-day-old hypocotyls of introgression lines (ILs), developed from the wild desert-adapted species *Solanum pennellii* and domesticated *Solanum lycopersicum* cv. M82 (Eshed and Zamir 1995), and found 5300 genes (~25% of total expressed genes) to have an eQTL, with roughly half in *cis* and half in *trans*. We also found this close to 50/50 ratio in the HP condition, whereas in the LN condition the ratio of *cis/trans* eQTLs was increased to 30/70. Research in yeast indicated that the detection of *trans*-acting eQTLs is more strongly affected by the power of the study than detection of *cis*-acting eQTLs (Albert et al. 2018). So, it is likely that in our study we would have even more *trans*-eQTLs relative to *cis*-eQTLs.

By comparing two different maternal environments in a population originating from two different genetic backgrounds, many different maternal environment-specific eQTLs were detected. This underlines the interplay between genetics and nutrient environment in our study. However, upon enrichment analysis of genes regulated by the environment-specific *trans*-eQTL hotspots, no obvious terms linking to the HP and LN environments were uncovered. Still, this information might be relevant for uncovering molecular mechanisms underlying the traits previously identified to co-locate with these *trans*-eQTL hotspots (Geshnizjani et al. 2020; Kazmi et al. 2012). Yet, we expect much of the variation caused by this interplay

will be uncovered in future studies increasing numbers of different timepoints, environments, and genotypes. It should be noted that, considering the relatively limited amount of RILs (49 RILs in the HP treatment and 52 RILs in the LN treatment in this study), eQTL mapping power can be likely improved by measuring more RIL genotypes.

Mapping eQTL can help in the identification of causal genes underlying phenotypic QTL. In our study, the resolution for identifying causal genes is mostly limited by the recombination frequency in the population used, which is limited in RILs with on-average two crossovers per chromosome. Next to recombination events, identification of causal polymorphic genes in this and other eQTL studies can be assisted by using prior knowledge (Hartanto et al. 2022), and more detailed data on the number and type of polymorphisms between tomato lines, such as frameshifts (Kevei et al. 2015) and copy number variations (Razali et al. 2018). Moreover, combining eQTLs with QTLs obtained using phenotypic trait data (Geshnizjani et al. 2019, 2020; Khan et al. 2012), as well as other molecular data such as proteomics and/or metabolomics (Kazmi et al. 2017), will contribute to obtaining mechanistic insight on how genotypic variation leads to phenotypic variation between individuals at a systemic level. Furthermore, these eQTLs could be used as a lead in studies with a larger source of wild genotypes and combined with GWAS (Bauchet et al. 2017; Chang et al. 2018; Mata-Nicolas et al. 2020; Ye et al. 2019; Zhang et al. 2015; Zhao et al. 2019), to pinpoint causal polymorphisms underlying variation at both the molecular and phenotypic levels.

Methods

Plant lines, growth conditions, and nutrient treatments

The mother plants (maternal conditions) were cultivated as described in Kazmi et al. (2012) and Geshnizjani et al. (2020) (Khan et al. 2012), in the greenhouse at Wageningen University, the Netherlands. In short, the parental lines *Solanum lycopersicum* cv. Money maker (MM) and *Solanum pimpinellifolium* accession CGN14498 (PI) as well as the derived recombinant inbred lines (RILs; (Voorrips et al. 2000); Supplementary Table 1) were grown on rockwool under standard nutrient conditions (14 mM nitrate and 1 mM phosphate) with a 16-h light (25 °C) and 8-h darkness (15 °C) photoperiod. From the moment the first flower opened, the plants were fertilized with the specific nutrient solutions, low nitrate (2.4 mM Nitrate, 1 mM Phosphate), and high phosphate (14 mM

Nitrate, 5 mM Phosphate) in two biological replicates per environment. The seeds were collected from healthy and fully ripened fruits, and the pulp still attached to the seeds was removed with 1% hydrochloric acid (HCl) and a mesh sieve (for 30 min). Water was used to remove the remaining HCl and pulp. For disinfection, seeds were treated with trisodium phosphate ($\text{Na}_3\text{PO}_4 \cdot 12\text{H}_2\text{O}$) for 15 min. Subsequently, seeds were dried at 20 °C for 3 days on a clean filter paper in ambient conditions. The seeds were then stored in paper bags at room temperature.

RNA-isolation, library prep, and RNA-seq

We used 10 mg grinded powder derived from 30 whole, dry, brushed, after-ripened seeds (12 months after harvest) of parental lines, and the RILs grown under different nutrient environments in a generalized genetical genomics design (Li et al. 2008, 2009b) to extract total RNA. For the HP treatment, 3 replicates for the parental lines and 49 unique RILs were sequenced (one replicate per RIL). For the LN treatment, 3 replicates for the parental lines and 52 unique RILs were sequenced (one replicate per RIL) (Supplementary Table 13). RNA was isolated using the NucleoSpin RNA plant isolation kit (Macherey–Nagel 740,949) with on-column DNA digestion and adding Plant RNA isolation Aid (Life technologies) according to the manufacturer’s protocol and instructions. Strand-specific RNA-seq libraries were prepared from each RNA sample using the TruSeq RNA kit from Illumina according to manufacturer’s instructions. Poly-A-selected mRNA was sequenced using the Illumina HiSeq2500 sequencer, producing strand-specific single-end reads of 100 nucleotides. Raw sequence reads can be found in the Sequence Read Archive (SRA; www.ncbi.nlm.nih.gov/sra) under ID: PRJNA704909.

Alignment and SNP calling

Reads were trimmed using Trimmomatic (version 0.33, (Bolger et al. 2014) to remove low-quality nucleotides. Trimmed reads were subsequently mapped to the Tomato SL4.0 reference genome with the ITAG4.0 annotation (Hosmani et al. 2020) using the HISAT2 software (version 2.1.0 (Kim et al. 2015) with the `-dta-cufflinks` option. The resulting SAM alignment files were sorted and indexed using samtools version 1.9 (Li et al. 2009a). SNPs were called using bcftools mpileup with a minimum read depth of 3.

Generation of a physical and genetic map from RNA-seq data

The physical map used for mapping the eQTLs was made from the RNA-seq data following the protocol described

in Serin et al. (2017) and Snoek et al. (2019). With the following modifications: SNPs were filtered for those that were consistently found in all replicates of the parental lines and observed in all RILs. Then, the genotype per RIL was determined per sliding bin of 100 SNPs where the mean position of those SNPs was taken as the physical position of the obtained marker.

The genetic map was constructed by converting the genotype probabilities to the most likely genotype and subsequently determining the number of recombinations between subsequent markers. The chance of recombination was used to generate a centimorgan (cM) map.

Quantification of RNAseq

Before mRNA abundance analysis, between 12 and 31 M reads per sample were mapped to the SL4.0 genome with ITAG4.0 annotation (Hosmani et al. 2020) using HISAT2 as described above. The mRNA abundance was quantified to counts using Stringtie (Pertea et al. 2015) with the options `-e`, `-B`, and `-G`. In R, the counts were used to calculate transcripts per million (TPM). The TPM values were \log_2 -transformed by

$$\text{TPM}_{\log} = \log_2(\text{TPM} + 1)$$

Additionally, to use for statistics, also a ratio with the average was calculated, by

$$\text{TPM}_{\text{rat},i,j} = \log_2\left(\frac{\text{TPM}_{i,j}}{\overline{\text{TPM}_j}}\right)$$

where the \log_2 was calculated for each transcript i of sample j by dividing over the average value for that transcript $\overline{\text{TPM}_j}$ over all samples j . After transformation, the transcripts were filtered for $\text{TPM}_{\log} > 0$ and detection in all samples.

mRNA abundance analysis and QTL analyses

The analyses reported below were conducted in “R” (version 3.5.3, $\times 64$) (R-Core-Team 2017) with custom written scripts, accessible via https://git.wur.nl/publicshed_papers/sterken_tomato-eqtl_2021. For analysis, the `dplyr` and `tidyr` packages were used for data organization (Wickham et al. 2018; Wickham 2018), and plots were generated using `ggplot2` (Wickham 2009).

Treatment-related mRNA abundance differences

The principal component analysis comparing the mRNA abundances was done on the TPM_{rat} -transformed data,

using the *prcomp* function in “R.” The mRNA abundance differences between treatments were tested between the LN and HP treatments using the linear model

$$\text{TMP}_{\log,i} = T_i + e_i$$

where $\text{TMP}_{\log,i}$ is the abundance level of transcript i (one of 14,772 transcripts) in RIL j ($n=55$ for the HP treatment and $n=58$ for the LN treatment; these include the RILs and the parental replicates per condition), T is the treatment (HP or LN), and e is the error term. To reduce the chance of detecting differences due to genetic variation, a strict multiple testing correction was applied (Bonferroni) using *p.adjust*. The threshold for significance was $-\log_{10}(p) > 5.47$ (FDR = 0.05).

To determine the effect of treatment on the differences in mRNA abundance between the parental lines, we ran a linear model explaining the differences due to treatment and line effects on the MM and PI parental data. The model used was

$$\text{TMP}_{\log,i,j} = T_{i,j} + L_{i,j} + T_{i,j} \times L_{i,j} + e_{i,j}$$

where $\text{TMP}_{\log,i,j}$ is the abundance level of transcript i (one of 14,772 transcripts) in parental replicate j ($n=3$ for both treatments for MM and PI), T is the treatment (HP or LN), L is the line (MM or PI), and e is the error term. Values were corrected for multiple testing using *p.adjust* following the Benjamini–Hochberg algorithm. The thresholds for FDR = 0.05 were: $-\log_{10}(p) = 1.71$ for line, $-\log_{10}(p) = 2.08$ for treatment, and $-\log_{10}(p) = 2.89$ for the interaction between line and treatment. We took the most stringent p value, $-\log_{10}(p) = 2.89$ as threshold to determine significance.

Transgression

Transgression was calculated by counting the number of lines with expression levels beyond three standard deviations from the mean of the parental lines (as in Brem and Kruglyak 2005); $\mu \pm 3\sigma$. This was done for both treatments separately. The lower boundary was established by the parental line with the lowest mean, and the upper boundary was established by the parental line with the highest mean. The standard deviation used to determine transgression (σ) was calculated as the pooled standard deviation of the two parental lines ($n=3$ for both).

Significance of the transgression was calculated by permutation. The expression values were randomized over the line designations and the same test as above was conducted. This was repeated 1000 times for each transcript, so the obtained values could be used as the by-chance distribution. The 50th highest value was used as the false discovery rate (FDR) = 0.05 threshold.

Heritability

The heritability was calculated by estimating the genotypic variance in the RILs and the remaining variance (e.g., measurement error) in the parental lines (as in Keurentjes et al. (2007)). This was done for both treatments separately, by

$$H_{\text{RIL}}^2 = \frac{V_{\text{RIL}} - V_e}{V_{\text{RIL}}}$$

where V_{RIL} is the variance within the RIL population and V_e is the pooled variance of both parental lines.

To establish whether the heritability was significant and not outlier driven, we applied a permutation approach [as in Vinuela et al. (2012)]. The trait values were randomized over the line designations and the heritability calculation were repeated. This was done 1000 times for each transcript to generate a by-chance distribution. The 50th highest value was used as the FDR = 0.05 threshold.

eQTL mapping

For eQTL mapping a single marker model was used and was applied separately for both treatments (as in (Snoek et al. 2017; Sterken et al. 2017)). QTLs were mapped using the model

$$\text{TPM}_{\log,i,j} = x_j + e_j$$

where $\text{TPM}_{\log,i,j}$ is the expression level of transcript i (one of 14,772 transcripts) in RIL j ($n=49$ for the HP treatment and $n=52$ for the LN treatment). The expression levels were explained over the genotype on marker location x ($x = 1, 2, \dots, 4515$) of RIL j .

To determine the reliability of the detected QTLs and correct for multiple testing, a permutation approach was used. As in the other permutations, the expression levels were randomly distributed over the lines and this randomized set was mapped again according to the procedure described above, which was repeated 10 times. To determine the FDR, we applied a correction for multiple testing under dependency (Benjamini and Yekutieli 2001)

$$\frac{\text{FDS}}{\text{RDS}} \leq \frac{m_0}{m} \times q \times \log(m)$$

where FDS (false discovery) is the number of eQTLs detected in the permutation and the RDS (real discovery) is the number of eQTLs detected in the QTL mapping at a specific significance level. The number of true null hypotheses tested (m_0) was $14,772 \cdot \text{RDS}$, where the number of hypotheses tested (m) was the number of transcripts,

14,772. The q -value was set at 0.05, which led to a threshold of $-\log_{10}(p) = 3.7$ for the LN treatment and $-\log_{10}(p) = 3.9$ for the HP treatment. To keep comparisons straightforward (similar effect sizes), analyses were conducted at the most stringent threshold ($-\log_{10}(p) > 3.9$).

The eQTL types (*cis* or *trans*) were called based on distance to the gene encoding the affected transcript. A *trans*-eQTL had to be located at least 1 Mb from the gene. Furthermore, we calculated the confidence interval of the QTL as a 1.5-drop from the highest $-\log_{10}(p)$. For a *trans*-eQTL to be called, the location of the affected transcript was required to be outside of this confidence interval as well.

Trans-band identification

Identification of regulatory hotspots (*trans*-bands) was based on assessing whether the number of *trans*-eQTLs mapped to a locus exceeded the expected number based on an equal genome-wide distribution [as in Rockman et al. 2010; Snoek et al. 2017]. We used a Poisson distribution to ascertain the significance of eQTL abundances per 2 Mb bin. For the HP treatment, we expected 15.8 *trans*-eQTL per bin, and for the LN treatment we expected 40.8 *trans*-eQTL per bin. We used a conservative threshold for calling a bin enriched in *trans*-eQTL, $p < 0.0001$. After identifying significant bins, adjacent bins (significant bins, with up to 1 non-significant bin in-between) were merged to a single *trans*-band.

Enrichment

GO enrichment was determined using the hypergeometric test in R on the GO annotation done for ITAG2.4 downloaded from AgriGO (www.bioinfo.cau.edu.cn/agriGO) (Tian et al. 2017) combined with the annotation for ITAG3.1 and expanded with the GO annotation of the Arabidopsis homologues. All expressed genes were used as background genes in the enrichment test.

Map and eQTL data in TomQTL

The physical map of the RIL population and the eQTL $-\log_{10}(p)$ -value scores are available for download and online exploration in TomQTL at <http://www.bioinformatics.nl/TomQTL/>, an interactive website based on AraQTL (Nijveen et al. 2017) and WormQTL2 (Snoek et al. 2020).

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00122-023-04322-0>.

Acknowledgements We thank Prof. Dr. Jan Kammenga of the Laboratory of Nematology of Wageningen University, Prof. Dr.

Dick de Ridder of the Laboratory of Bioinformatics of Wageningen University, and Prof. Dr. Berend Snel of the Theoretical Biology and Bioinformatics department of Utrecht University for their support.

Author Contribution statement WL and HWMH conceived the study, JMJJ, NG, LAJW, and JR performed the experiments, MGS, HN, and LBS analyzed and visualized the data, MGS and LBS wrote the manuscript with input from MvZ, HN, JMJJ, and WL. All authors approved the final version of this manuscript.

Funding This work was supported by Technology Foundation (STW), which is part of the Netherlands Organization for Scientific Research (NWO) (LW, JR, HN, WL). M.G.S. was supported by NWO domain Applied and Engineering Sciences VENI grant (17282).

Data Availability statement The datasets generated and analyzed during the current study are available in the Sequence Read Archive (SRA; www.ncbi.nlm.nih.gov/sra) repository, under ID: PRJNA704909.

Declarations

Conflict of interest The authors have not disclosed any competing interests.

One sentence summary The maternal nutritional environment has a clear influence on the expression of genes in mature dry tomato seeds and the loci affecting this expression shed light on regulation of seed quality.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Albert E, Segura V, Gricourt J, Bonnefoi J, Derivot L, Causse M (2016) Association mapping reveals the genetic architecture of tomato response to water deficit: focus on major fruit quality traits. *J Exp Bot* 67:6413–6430
- Albert FW, Bloom JS, Siegel J, Day L, Kruglyak L (2018) Genetics of trans-regulatory variation in gene expression. *eLife* 7:e35471
- Alboresi A, Gestin C, Leydecker MT, Bedu M, Meyer C, Truong HN (2005) Nitrate, a signal relieving seed dormancy in Arabidopsis. *Plant Cell Environ* 28:500–512
- Bauchet G, Munos S, Sauvage C, Bonnet J, Grivet L, Causse M (2014) Genes involved in floral meristem in tomato exhibit drastically reduced genetic diversity and signature of selection. *BMC Plant Biol* 14:279
- Bauchet G, Grenier S, Samson N, Segura V, Kende A, Beekwilder J, Cankar K, Gallois JL, Gricourt J, Bonnet J, Baxter C, Grivet L, Causse M (2017) Identification of major loci and genomic regions

- controlling acid and volatile content in tomato fruit: implications for flavor improvement. *New Phytol* 215:624–641
- Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *Ann Stat* 29:23
- Blanca J, Montero-Pau J, Sauvage C, Bauchet G, Illa E, Diez MJ, Francis D, Causse M, van der Knaap E, Canizares J (2015) Genomic variation in tomato, from wild ancestors to contemporary breeding accessions. *BMC Genom* 16:257
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120
- Brem RB, Kruglyak L (2005) The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proc Natl Acad Sci USA* 102:1572–1577
- Capel C, Fernandez del Carmen A, Alba JM, Lima-Silva V, Hernandez-Gras F, Salinas M, Boronat A, Angosto T, Botella MA, Fernandez-Munoz R, Granell A, Capel J, Lozano R (2015) Wide-genome QTL mapping of fruit quality traits in a tomato RIL population derived from the wild-relative species *Solanum pimpinellifolium* L. *Theoretische Und Angewandte Genetik* 128:2019–2035
- Capel C, Yuste-Lisbona FJ, Lopez-Casado G, Angosto T, Cuartero J, Lozano R, Capel J (2017) Multi-environment QTL mapping reveals genetic architecture of fruit cracking in a tomato RIL *Solanum lycopersicum* × *S. pimpinellifolium* population. *Theoretische Und Angewandte Genetik* 130:213–222
- Celik I, Gurbuz N, Uncu AT, Frary A, Doganlar S (2017) Genome-wide SNP discovery and QTL mapping for fruit quality traits in inbred backcross lines (IBLs) of *Solanum pimpinellifolium* using genotyping by sequencing. *BMC Genom* 18:1
- Chang J, Yu T, Yang Q, Li C, Xiong C, Gao S, Xie Q, Zheng F, Li H, Tian Z, Yang C, Ye Z (2018) Hair, encoding a single C2H2 zinc-finger protein, regulates multicellular trichome formation in tomato. *Plant J: Cell Mol Biol* 96:90–102
- Cubillos FA, Stegle O, Grondin C, Canut M, Tisne S, Gy I, Loudet O (2014) Extensive cis-regulatory variation robust to environmental perturbation in Arabidopsis. *Plant Cell* 26:4298–4310
- de Souza VD, Willems L, van Arkel J, Dekkers BJW, Hilhorst HWM, Bentsink L (2016) Galactinol as marker for seed longevity. *Plant Sci Int J Exp Plant Biol* 246:112–118
- Delouche JC (1980) Environmental effects on seed development and seed quality. *HortScience* 15:775–780
- Delouche J, Baskin C (1971) Determinants of seed quality. *Short Course Seedsmen* 14:53–68
- Demir I, Mavi K, Oztokan C (2004) Changes in germination and potential longevity of watermelon (*Citrullus lanatus*) seeds during development. *N Z J Crop Hortic Sci* 32:139–145
- Demirci S, van Dijk AD, Sanchez Perez G, Aflitos SA, de Ridder D, Peters SA (2017) Distribution, position and genomic characteristics of crossovers in tomato recombinant inbred lines derived from an interspecific cross between *Solanum lycopersicum* and *Solanum pimpinellifolium*. *Plant J Cell Mol Biol* 89:554–564
- Doebley JF, Gaut BS, Smith BD (2006) The molecular genetics of crop domestication. *Cell* 127:1309–1321
- Donohue K (2009) Completing the cycle: maternal effects as the missing link in plant life histories. *Philos Trans R Soc Lond b: Biol Sci* 364:1059–1074
- Duermeyer L, Khodapanahi E, Yan D, Krapp A, Rothstein S, Nambara E (2018) Regulation of seed dormancy and germination by nitrate. *Seed Sci Res* 28:150–157
- 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* 141:1147–1162
- Geshnizjani N, Ghaderi-Far F, Willems LAJ, Hilhorst HWM, Ligterink W (2018) Characterization of and genetic variation for tomato seed thermo-inhibition and thermo-dormancy. *BMC Plant Biol* 18:229
- Geshnizjani N, Sarikhani Khorami S, Willems LAJ, Snoek BL, Hilhorst HWM, Ligterink W (2019) The interaction between genotype and maternal nutritional environments affects tomato seed and seedling quality. *J Exp Bot* 70:2905–2918
- Geshnizjani N, Snoek BL, Willems LAJ, Rienstra JA, Nijveen H, Hilhorst HWM, Ligterink W (2020) Detection of QTLs for genotype × environment interactions in tomato seeds and seedlings. *Plant Cell Environ* 43:1973–1988
- Giovannoni J (2001) Molecular biology of fruit maturation and ripening. *Annu Rev Plant Biol* 52:725–749
- Hartanto M, Joosen RVL, Snoek BL, Willems LAJ, Sterken MG, de Ridder D, Hilhorst HWM, Ligterink W, Nijveen H (2020) Network analysis prioritizes DEWAX and ICE1 as the candidate genes for two major eQTL hotspots in seed germination. *G3 Genes Genomes. Genetics* 10:4215–4226
- Hartanto M, Sami AA, de Ridder D, Nijveen H (2022) Prioritizing candidate eQTL causal genes in Arabidopsis using random forests. *G3 Genes Genomes. Genetics* 12:jkac255
- He H, de Souza VD, Snoek LB, Schnabel S, Nijveen H, Hilhorst H, Bentsink L (2014) Interaction between parental environment and genotype affects plant and seed performance in Arabidopsis. *J Exp Bot* 65:6603–6615
- He R, Yu D, Li X, Duan G, Zhang Y, Tang D, Zhao X, Liu X (2016) F-box gene FOA2 regulates GA- and ABA-mediated seed germination in Arabidopsis. *Sci China Life Sci* 59:1192–1194
- Hosmani PS, Mirella F-G, van de Geest H, Maumus F, Bakker LV, Schijlen E, van Haarst J, Cordewener J, Sanchez-Perez G, Peters S, Fei Z, Giovannoni JJ, Mueller LA, Saha S (2020) An improved de novo assembly and annotation of the tomato reference genome using single-molecule sequencing, Hi-C proximity ligation and optical maps. *BioRxiv*
- Jansen RC, Nap JP (2001) Genetical genomics: the added value from segregation. *Trends Genet* 17:388–391
- Jimenez-Gomez JM, Wallace AD, Maloof JN (2010) Network analysis identifies ELF3 as a QTL for the shade avoidance response in Arabidopsis. *PLoS Genet* 6:e1001100
- Joosen RV, Arends D, Willems LA, Ligterink W, Jansen RC, Hilhorst HW (2012) Visualizing the genetic landscape of Arabidopsis seed performance. *Plant Physiol* 158:570–589
- Kawakatsu T, Huang SS, Jupe F, Sasaki E, Schmitz RJ, Urich MA, Castanon R, Nery JR, Barragan C, He Y, Chen H, Dubin M, Lee CR, Wang C, Bemm F, Becker C, O'Neil R, O'Malley RC, Quarless DX, Genomes C, Schork NJ, Weigel D, Nordborg M, Ecker JR (2016) Epigenomic diversity in a global collection of *Arabidopsis thaliana* accessions. *Cell* 166:492–505
- Kazmi RH, Khan N, Willems LA, Van Heusden AW, 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* × *Solanum pimpinellifolium*. *Plant Cell Environ* 35:929–951
- Kazmi RH, Willems LAJ, Joosen RVL, Khan N, Ligterink W, Hilhorst HWM (2017) Metabolomic analysis of tomato seed germination. *Metab: off J Metabol Soc* 13:145
- Keurentjes JJ, Fu J, Terpstra IR, Garcia JM, van den Ackerveken G, Snoek LB, Peeters AJ, Vreugdenhil D, Koornneef M, Jansen RC (2007) Regulatory network construction in Arabidopsis by using genome-wide gene expression quantitative trait loci. *Proc Natl Acad Sci USA* 104:1708–1713
- Kevei Z, King RC, Mohareb F, Sergeant MJ, Awan SZ, Thompson AJ (2015) Resequencing at >=40-fold depth of the parental genomes of a *Solanum lycopersicum* × *S. pimpinellifolium* recombinant inbred line population and characterization of frame-shift InDels that are highly likely to perturb protein function. *G3 Genes Genomes Genet* 5:971–981

- Khan N, Kazmi RH, Willems LA, van Heusden AW, Ligterink W, Hilhorst HW (2012) Exploring the natural variation for seedling traits and their link with seed dimensions in tomato. *PLoS ONE* 7:e43991
- Kim D, Langmead B, Salzberg SL (2015) HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* 12:357–360
- Li Y, Alvarez OA, Gutteling EW, Tijsterman M, Fu J, Riksen JA, Hazendonk E, Prins P, Plasterk RH, Jansen RC, Breitling R, Kammenga JE (2006) Mapping determinants of gene expression plasticity by genetical genomics in *C. elegans*. *PLoS Genet* 2:e 222
- Li Y, Breitling R, Jansen RC (2008) Generalizing genetical genomics: getting added value from environmental perturbation. *Trends Genet* 24:518–524
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009a) The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079
- Li Y, Swertz MA, Vera G, Fu J, Breitling R, Jansen RC (2009b) designGG: an R-package and web tool for the optimal design of genetical genomics experiments. *BMC Bioinform* 10:188
- Lin T, Zhu G, Zhang J, Xu X, Yu Q, Zheng Z, Zhang Z, Lun Y, Li S, Wang X, Huang Z, Li J, Zhang C, Wang T, Zhang Y, Wang A, Zhang Y, Lin K, Li C, Xiong G, Xue Y, Mazzucato A, Causse M, Fei Z, Giovannoni JJ, Chetelat RT, Zamir D, Stadler T, Li J, Ye Z, Du Y, Huang S (2014) Genomic analyses provide insights into the history of tomato breeding. *Nat Genet* 46:1220–1226
- Matakadiadis T, Alboresi A, Jikumaru Y, Tatematsu K, Pichon O, Renou JP, Kamiya Y, Nambara E, Truong HN (2009) The Arabidopsis abscisic acid catabolic gene CYP707A2 plays a key role in nitrate control of seed dormancy. *Plant Physiol* 149:949–960
- Mata-Nicolas E, Montero-Pau J, Gimeno-Paez E, Garcia-Carpintero V, Ziarsolo P, Menda N, Mueller LA, Blanca J, Canizares J, van der Knaap E, Diez MJ (2020) Exploiting the diversity of tomato: the development of a phenotypically and genetically detailed germplasm collection. *Horticult Res* 7:66
- McCouch S (2004) Diversifying selection in plant breeding. *PLoS Biol* 2:e347
- Munir J, Dorn L, Donohue K, Schmitt J (2001) The influence of maternal photoperiod on germination requirements in *Arabidopsis thaliana*. *Am J Bot* 88:1240–1249
- Nijveen H, Ligterink W, Keurentjes JJ, Loudet O, Long J, Sterken MG, Prins P, Hilhorst HW, de Ridder D, Kammenga JE, Snoek BL (2017) AraQTL—workbench and archive for systems genetics in *Arabidopsis thaliana*. *Plant J Cell Mol Biol* 89:1225–1235
- Oyserman BO, Flores SS, Griffioen T, Pan X, van der Wijk E, Pronk L, Lokhorst W, Nurfikari A, Paulson J, Movassagh M, Stopnisek N, Kupczok A, Cordovez V, Carrión VJ, Ligterink W, Snoek BL, Medema MH, Raaijmakers JM (2022) Disentangling the genetic basis of rhizosphere microbiome assembly in tomato. *Nature communications* 13:3228
- Pascual L, Albert E, Sauvage C, Duangjit J, Bouchet JP, Bitton F, Desplat N, Brunel D, Le Paslier MC, Ranc N, Bruguier L, Chauchard B, Verschave P, Causse M (2016) Dissecting quantitative trait variation in the resequencing era: complementarity of bi-parental, multi-parental and association panels. *Plant Sci Int J Exp Plant Biol* 242:120–130
- Perteau M, Perteau GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* 33:290–295
- Pourrat Y, Jacques R (1975) The influence of photoperiodic conditions received by the mother plant on morphological and physiological characteristics of *Chenopodium polyspermum* L. seeds. *Plant Sci Lett* 4:273–279
- Ranjan A, Budke JM, Rowland SD, Chitwood DH, Kumar R, Carriedo L, Ichihashi Y, Zumstein K, Maloof JN, Sinha NR (2016) eQTL regulating transcript levels associated with diverse biological processes in tomato. *Plant Physiol* 172:328–340
- Razali R, Bougouffa S, Morton MJL, Lightfoot DJ, Alam I, Essack M, Arold ST, Kamau AA, Schmockel SM, Pailles Y, Shahid M, Mitchell CT, Al-Babili S, Ho YS, Tester M, Bajic VB, Negrao S (2018) The genome sequence of the wild tomato *Solanum pimpinellifolium* provides insights into salinity tolerance. *Front Plant Sci* 9:1402
- Razifard H, Ramos A, Della Valle AL, Bodary C, Goetz E, Manser EJ, Li X, Zhang L, Visa S, Tieman D, van der Knaap E, Caicedo AL (2020) Genomic evidence for complex domestication history of the cultivated tomato in Latin America. *Mol Biol Evol* 37:1118–1132
- R-Core-Team (2017) R: a language and environment for statistical computing. R Foundation for Statistical Computing Vienna, Austria
- Rockman MV, Skrovaneck SS, Kruglyak L (2010) Selection at linked sites shapes heritable phenotypic variation in *C. elegans*. *Science* 330:372–376
- Rothan C, Diouf I, Causse M (2019) Trait discovery and editing in tomato. *Plant J Cell Mol Biol* 97:73–90
- Rowse H, Finch-Savage W (2003) Hydrothermal threshold models can describe the germination response of carrot (*Daucus carota*) and onion (*Allium cepa*) seed populations across both sub-and supra-optimal temperatures. *New Phytol* 158:101–108
- Sauvage C, Segura V, Bauchet G, Stevens R, Do PT, Nikoloski Z, Fernie AR, Causse M (2014) Genome-wide association in tomato reveals 44 candidate loci for fruit metabolic traits. *Plant Physiol* 165:1120–1132
- Schachtman DP, Reid RJ, Ayling SM (1998) Phosphorus uptake by plants: from soil to cell. *Plant Physiol* 116:447–453
- 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. *Nat Biotechnol* 24:447–454
- Schmuths H, Bachmann K, Weber WE, Horres R, Hoffmann MH (2006) Effects of preconditioning and temperature during germination of 73 natural accessions of *Arabidopsis thaliana*. *Ann Bot* 97:623–634
- Serin EAR, Snoek LB, Nijveen H, Willems LAJ, Jimenez-Gomez JM, Hilhorst HWM, Ligterink W (2017) Construction of a high-density genetic map from RNA-Seq data for an Arabidopsis Bay-0 × Shahdara RIL population. *Front Genet* 8:201
- Snoek LB, Terpstra IR, Dekter R, Van den Ackerveken G, Peeters AJ (2012) Genetical genomics reveals large scale genotype-by-environment interactions in *Arabidopsis thaliana*. *Front Genet* 3:317
- Snoek BL, Sterken MG, Bevers RPJ, Volkens RJM, Van't Hof A, Brenchley R, Riksen JAG, Cossins A, Kammenga JE (2017) Contribution of trans regulatory eQTL to cryptic genetic variation in *C. elegans*. *BMC Genomics* 18:500
- Snoek BL, Volkens RJM, Nijveen H, Petersen C, Dirksen P, Sterken MG, Nakad R, Riksen JAG, Rosenstiel P, Stastna JJ, Braeckman BP, Harvey SC, Schulenburg H, Kammenga JE (2019) A multi-parent recombinant inbred line population of *C. elegans* allows identification of novel QTLs for complex life history traits. *BMC Biol* 17:24
- Snoek BL, Sterken MG, Hartanto M, van Zuilichem AJ, Kammenga JE, de Ridder D, Nijveen H (2020) WormQTL2: an interactive platform for systems genetics in *Caenorhabditis elegans*. *Database J Biol Databases Curation* 2020
- de Souza Vidigal D, He H, Hilhorst HWM, Willems LAJ, Bentsink L (2020) Arabidopsis in the wild—the effect of seasons on seed performance. *Plants* 9:576
- Sterken MG, van Bemmelen, van der Plaats L, Riksen JAG, Rodriguez M, Schmid T, Hajnal A, Kammenga JE, Snoek BL (2017) Ras/MAPK modifier loci revealed by eQTL in *Caenorhabditis elegans*. *G3 Genes Genomes Genet* 7:3185–3193

- Sterken MG, Bevers RPJ, Volkens RJM, Riksen JAG, Kammenga JE, Snoek BL (2019) Dissecting the eQTL micro-architecture in *Caenorhabditis elegans*. *BioRxiv*
- Terpstra IR, Snoek LB, Keurentjes JJ, Peeters AJ, van den Ackerveken G (2010) Regulatory network identification by genetical genomics: signaling downstream of the Arabidopsis receptor-like kinase ERECTA. *Plant Physiol* 154:1067–1078
- Tian T, Liu Y, Yan H, You Q, Yi X, Du Z, Xu W, Su Z (2017) agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res* 45:W122–W129
- Tomato Genome C (2012) The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485:635–641
- Tomato Genome Sequencing C, Aflitos S, Schijlen E, de Jong H, de Ridder D, Smit S, Finkers R, Wang J, Zhang G, Li N, Mao L, Bakker F, Dirks R, Breit T, Gravendeel B, Huits H, Struss D, Swanson-Wagner R, van Leeuwen H, van Ham RC, Fito L, Guignier L, Sevilla M, Ellul P, Ganko E, Kapur A, Reclus E, de Geus B, van de Geest H, Te Lintel HB, van Haarst J, Smits L, Koops A, Sanchez-Perez G, van Heusden AW, Visser R, Quan Z, Min J, Liao L, Wang X, Wang G, Yue Z, Yang X, Xu N, Schranz E, Smets E, Vos R, Rauwerda J, Ursem R, Schuit C, Kerns M, van den Berg J, Vriezen W, Janssen A, Datema E, Jahrman T, Moquet F, Bonnet J, Peters S (2014) Exploring genetic variation in the tomato (*Solanum section Lycopersicon*) clade by whole-genome sequencing. *Plant J Cell Mol Biol* 80:136–148
- Urbanczyk-Wochniak E, Fernie AR (2005) Metabolic profiling reveals altered nitrogen nutrient regimes have diverse effects on the metabolism of hydroponically-grown tomato (*Solanum lycopersicum*) plants. *J Exp Bot* 56:309–321
- Vinuela A, Snoek LB, Riksen JA, Kammenga JE (2010) Genome-wide gene expression regulation as a function of genotype and age in *C. elegans*. *Genome Res* 20:929–937
- Vinuela A, Snoek LB, Riksen JA, Kammenga JE (2012) Aging uncouples heritability and expression-QTL in *Caenorhabditis elegans*. *G3 Genes Genomes Genet* 2:597–605
- Viquez-Zamora M, Caro M, Finkers R, Tikunov Y, Bovy A, Visser RG, Bai Y, van Heusden S (2014) Mapping in the era of sequencing: high density genotyping and its application for mapping TYLCV resistance in *Solanum pimpinellifolium*. *BMC Genomics* 15:1152
- Voorrips RE, Verkerke W, Finkers R, Jongerius R, Kanne J (2000) Inheritance of taste components in tomato. *Acta Physiologicae Plantarum* 22:259–261
- Vosman B, Van't Westende WPC, Henken B, van Eekelen H, de Vos RCH, Voorrips RE (2018) Broad spectrum insect resistance and metabolites in close relatives of the cultivated tomato. *Euphytica Neth J Plant Breed* 214:46
- Wang Z, Hong Y, Zhu G, Li Y, Niu Q, Yao J, Hua K, Bai J, Zhu Y, Shi H, Huang S, Zhu JK (2020) Loss of salt tolerance during tomato domestication conferred by variation in a Na(+)/K(+) transporter. *EMBO J* 39:e103256
- West MA, Kim K, Kliebenstein DJ, van Leeuwen H, Michelmore RW, Doerge RW, St Clair DA (2007) Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in Arabidopsis. *Genetics* 175:1441–1450
- Wickham H (2009) *GGplot2: elegant graphics for data analysis*. Springer, New York, NY, USA
- Wickham H, Henry L (2018) *tidyr: Easily Tidy Data with 'spread()' and 'gather()' Functions*. R package version 0.8.1. <https://CRAN.R-project.org/package=tidyr>
- Wickham HF, François R, Henry L, Müller K (2018) *dplyr: a Grammar of Data Manipulation*. R package version 0.7.6. <http://CRAN.R-project.org/package=dplyr>
- Yang J, Wang Y, Shen H, Yang W (2014) In silico identification and experimental validation of insertion-deletion polymorphisms in tomato genome. *DNA Res Int J Rapid Public Rep Genes Genomes* 21:429–438
- Ye J, Li W, Ai G, Li C, Liu G, Chen W, Wang B, Wang W, Lu Y, Zhang J, Li H, Ouyang B, Zhang H, Fei Z, Giovannoni JJ, Ye Z, Zhang Y (2019) Genome-wide association analysis identifies a natural variation in basic helix-loop-helix transcription factor regulating ascorbate biosynthesis via D-mannose/L-galactose pathway in tomato. *PLoS Genet* 15:e1008149
- Zhang J, Zhao J, Xu Y, Liang J, Chang P, Yan F, Li M, Liang Y, Zou Z (2015) Genome-wide association mapping for tomato volatiles positively contributing to tomato flavor. *Front Plant Sci* 6:1042
- Zhang S, Yu H, Wang K, Zheng Z, Liu L, Xu M, Jiao Z, Li R, Liu X, Li J, Cui X (2018) Detection of major loci associated with the variation of 18 important agronomic traits between *Solanum pimpinellifolium* and cultivated tomatoes. *Plant J Cell Mol Biol* 95:312–323
- Zhao J, Sauvage C, Zhao J, Bitton F, Bauchet G, Liu D, Huang S, Tiesman DM, Klee HJ, Causse M (2019) Meta-analysis of genome-wide association studies provides insights into genetic control of tomato flavor. *Nat Commun* 10:1534

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.