Comparative QTL analysis in Arabidopsis and Tomato

Comparative QTL analysis in Arabidopsis and Tomato

Master Minor thesis

Zhou, Heming

951107987050

Master of Plant Sciences

Bioinformatics department

BIF-80324

Nijveen, Harm; Hartanto, Margi

24th, Aug 2020

Wageningen University & Research

Wageningen, Netherlands

Abstract

Understanding genes directly related to production traits like seed germination has been increasingly crucial for tomato research. Meanwhile, research concerning with Arabidopsis has created sufficient data to investigate and to infer homologous genes in other species. Our experiment aimed to combine these two species to find candidate genes in tomato quickly. We proposed a cross-species QTL comparison based on similar traits. By comparing existing tomato QTL data with Arabidopsis, we built orthologous groups for QTL's underlying genes using computational biology approach. We collected causal genes in Arabidopsis seed germination and tried to find corresponding homologs in tomato. These homologs are selected as tomato candidate genes involved in seed germination after investigating orthologous gene databases and permutation test. Our methods provided candidates for future tomato research directions and a novel way to study QTL and their underlying genes.

Keywords: causal genes, orthologous groups, QTL, seed germination, tomato.

Table of contents

Abstract II

Introduction 1

Materials & Methods 2

QTL data 2

Construction of orthologous groups 3

Sample genes and comparison 3

Manual validation 4

Permutation test 5

Results 5

Reference orthogroups 5

QTL regions and causal genes mapping 10

Permutation test 11

Discussion 12

Reference 15

Introduction

The future global warming and drastic climate changes pose a threat to tomato growing. Especially heat and salt stress during seed germination and seedling stages (Geshnizjani et al., 2019). These traits including seed germination and quality are highly correlated with agricultural production (Khan et al., 2012). Many of these traits are quantitative. Quantitative traits are determined by accumulating complexed genetic and environmental effects (Mackay, 2001). Quantitative trait loci (QTL) analysis is a common practice in analyzing complex phenotypes that attempts to explain genetic effects on variations of complex traits using statistical methods (Falconer, 1996; Kearsey, 1998). Understanding how to improve crop production and resistance against all kinds of stress is essential because the growing demands for foods and deteriorating environments require faster, more accurate methods to improve crops (Lin et al., 2019; Meadows et al., 1992). Currently, the common methods require long time cycle of growing crops, constructing needed populations and efforts on phenotyping. QTL analysis usually cannot give the responsible genes directly, in most cases it shows intervals on chromosomes that are linked with the trait variation (Miles & Wayne, 2008).

Numerous high-density genetic maps, increasingly complete gene orthologous databases, and plenty of explanation on genetic pathways have greatly enhanced our understanding on Arabidopsis (Berardini et al., 2015; Nunes-Nesi et al., 2019; Elise A. R. Serin et al., 2017; Yao et al., 2011). Previous physiological and genetic analysis has explained the mechanism of controlling seed germination and dormancy in Arabidopsis. These results have made Arabidopsis well understood as a model plant. Meanwhile tomato, as one of the most crucial crops in the world, is regarded as a future model plant for high production and resistance research (Geshnizjani et al., 2020; Sato et al., 2012). Previous experiments offered many outcomes on environmental effects combined with genetic ones on tomato seedling (Geshnizjani et al., 2019, 2020). However, the current tomato research has not given enough information on the underlying responsible genes of seed germination compared to Arabidopsis. Comparative genomics analysis has been popular in detecting tomato resistance for diseases over these years (Cui et al., 2017). Yet, it is less popular to compare tomato seed germination with other species to much extent. Hence, combining Arabidopsis and tomato is a new trend to maximize exploiting existing data to find more results.

Research nowadays requires more innovative methods that accurately detect underlying genes for complex traits as seed germination (Kazmi et al., 2012; E. A. R. Serin et al., 2017). Although scientists have been using linkage mapping to determine trait controlling loci, these loci can contain hundreds of underlying candidate genes and thus, require researchers more time and intensive mapping to find the causal genes (Lin et al., 2019). Causal genes refer to variation within genes' sequences and states that directly cause phenotypic variation (Hormozdiari et al., 2015; Weirauch et al., 2014). Therefore, to identify these causal genes and their function becomes important for improving these traits. QTL linkage mapping and genome-wide association study (GWAS) are popular methods to determine QTL and causal genes. Both methods have contributed greatly to identifying loci and causal genes. But their risks and requirements such as time-consuming and needs to construct certain populations have left room for improvement (Lin et al., 2019; Rueedi et al., 2014).

Here we present a new-fashioned method: Cross-species QTL comparison analysis. To cope with stress conditions, many plants keep a similar mechanism during evolution. Similar trait under same conditions may inherit similar molecular regulatory mechanism between different plants. For example, lettuce holds similar abscisic acid (ABA) and gibberellic acid (GA) molecular pathways as Arabidopsis during seed dormancy (Argyris et al., 2008; Huo & Bradford, 2015). Although Arabidopsis and tomato are distantly related plants, we believed some functions are conserved and similar in both dicot species especially when it comes to abiotic stress response for seeds. The QTL mapping data of both species are prepared to conduct such analysis (Kazmi et al., 2012; E. A. R. Serin et al., 2017). Therefore, we selected genes derived from these QTL regions and tried to find the orthologous links among them. In this project, we primarily focused on causal genes in relation to it controlling seed germination trait. We collected causal genes from *Arabidopsis thaliana*, a well-studied model plant, to detect the potential counterparts in tomato. We hoped this method can shorten the time needed for tomato resistance research and be promoted to other species in the future.

Materials & Methods

Our *in-silico* experiment made use of existing and relatively complete seed germination QTL data, and well-performed tool for orthologs inference. We collected and summarized the existing data at first. Then we designed functions to initiate testing the hypothesis. Furthermore, we verified our results by running permutation and manually examining various online sources.

QTL data

We collected Arabidopsis seed germination related QTL data from the Arabidopsis seed performance research by Serin et al. (2017). The data is from A. thaliana accessions Bay-0 and Sha and a Bay-0 × Sha recombinant inbred line (RIL). The corresponding data of tomato were taken from tomato research Kazmi et al. (2012). The tomato QTL analysis was derived from a RIL of *S. lycopersicum* (cultivar Moneymaker) x *S. pimpinellifolium* (Voorrips et al., 2000). The common seed germination traits include maximum germination capacity (Gmax), rate of germination (t_{10}, t_{50}), uniformity of germination (U8416, U7525), mean germination rate (MGR), and area under the germination curve (AUC) etc. These data have a common structure which includes trait names, stress conditions, chromosome numbers, nearest marker peak position, LOD scores. Nearest marker is a representative of QTL position on chromosomes. LOD score, or logarithm of odds, is a statistical evidence of detecting QTL presence (Churchill & Doerge, 1994). According to Lander & Botstein (1989), LOD score should be above 2 or 3 to guarantee low false positive for QTL detection. We chose the QTL whose LOD >2 and filtered out the QTL data that had no marker peak position for reliability.

Conditions	Trait	Chr	Marker Peak	LOD
Cold Stress (12 °C)	AUC	1	69227784	2.4
Cold Stress (12 °C)	Gmax	1	69227784	2.07
High Temperature I (35 °C)	MGR	1	69227784	2.79
Salt I(-0.3MPa NaCl)	MGR	1	7044030	2.02
Control	AUC	2	34914156	2.57

Table 1. Example of tomato seed germination QTL from research (Kazmi et al., 2012).

For searching genes located in a QTL region, we needed to select a reference genome. We used reference genome data release TAIR 10 for *Arabidopsis thaliana* and ITG version 2.5 reference genome Sol 2.5 from *'Solgenomics'* network for *Solanum Lycopersicum* (Berardini et al., 2015; Fernandez-Pozo et al., 2014). Both genomes were close to the time of QTL data in terms of the number of annotated genes and genes' position. First, we extracted only mRNA entries for gene references and remove all the rest entries. Second, we removed the last digital number at the end of gene's id (e.g. 'Solyc03g112280.2.1' to 'Solyc03g112280.2'). This is for removing duplicated entries and simplifying later online database investigation. After filtering redundant entries, the total tomato reference genome has 33,820 mRNA, and Arabidopsis has 27,628. These genes were searched on '*Solgenomics'* network (Fernandez-Pozo et al., 2014) and TAIR database (Berardini et al., 2015) to download their encoding protein sequences for later orthologs comparison. From here we refer to protein-encoding mRNA as gene name.

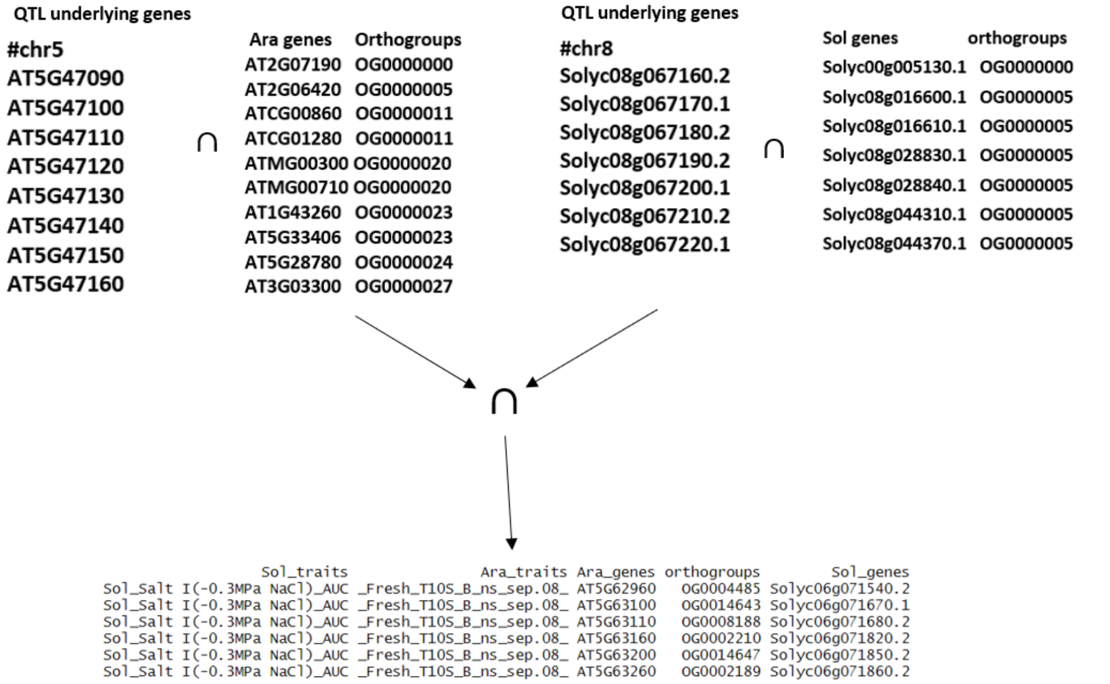
Construction of orthologous groups

We employed OrthoFinder (Emms & Kelly, 2019) to construct reference orthologous groups (orthogroups) based on all collected protein sequences from Arabidopsis and tomato in FASTA form. Through its state-of-the-art algorithm, OrthoFinder can yield orthogroups that contain a list of gene names in the same group based on phylogenetic inference. We only selected groups that contain genes from both species. The rest of the groups were regarded as paralogous groups and not taken into consideration. Besides OrthoFinder, the remaining experiment was done in R (R Core Team, 2018).

Sample genes and comparison

For each QTL to begin with, we extracted genome file with corresponding chromosome numbers that matched with QTL's chromosome positions and sampled 1Mb long at both sides of the marker peak as an interval on the genome. If the peak is close to the boundary of the chromosome and 1Mb length exceeded, we sampled a narrower or asymmetric interval based on the marker peak's position. The lists of genes within intervals were collected and attached with trait name in a single file and awaited to be compared.

From there on we started cross-all traits comparison for searching orthologs in both species. Each list of gene names first intersected with reference orthogroups. If gene names matched entries in the reference orthogroups, then these entries were collected in a table alongside with their homologs in the other species. Two tables then intercrossed and recorded the gene names which were shared by traits from Arabidopsis and tomato. The recorded gene names were attached with traits names of both species.



QTL underlying genes

#chr5
AT5G47090
AT5G47100
AT5G47110
AT5G47120
AT5G47130
AT5G47140
AT5G47150
AT5G47160

Ara genes	Orthogroups
AT2G07190	OG0000000
AT2G06420	OG0000005
ATCG00860	OG0000011
ATCG01280	OG0000011
ATMG00300	OG0000020
ATMG00710	OG0000020
AT1G43260	OG0000023
AT5G33406	OG0000023
AT5G28780	OG0000024
AT3G03300	OG0000027

QTL underlying genes

#chr8
Solyc08g067160.2
Solyc08g067170.1
Solyc08g067180.2
Solyc08g067190.2
Solyc08g067200.1
Solyc08g067210.2
Solyc08g067220.1

Sol genes	orthogroups
Solyc00g005130.1	OG0000000
Solyc08g016600.1	OG0000005
Solyc08g016610.1	OG0000005
Solyc08g028830.1	OG0000005
Solyc08g028840.1	OG0000005
Solyc08g044310.1	OG0000005
Solyc08g044370.1	OG0000005

```
                  Sol_traits                Ara_traits Ara_genes orthogroups        Sol_genes
Sol_Salt I(-0.3MPa NaCl)_AUC _Fresh_T10S_B_ns_sep.08_ AT5G62960   OG0004485 Solyc06g071540.2
Sol_Salt I(-0.3MPa NaCl)_AUC _Fresh_T10S_B_ns_sep.08_ AT5G63100   OG0014643 Solyc06g071670.1
Sol_Salt I(-0.3MPa NaCl)_AUC _Fresh_T10S_B_ns_sep.08_ AT5G63110   OG0008188 Solyc06g071680.2
Sol_Salt I(-0.3MPa NaCl)_AUC _Fresh_T10S_B_ns_sep.08_ AT5G63160   OG0002210 Solyc06g071820.2
Sol_Salt I(-0.3MPa NaCl)_AUC _Fresh_T10S_B_ns_sep.08_ AT5G63200   OG0014647 Solyc06g071850.2
Sol_Salt I(-0.3MPa NaCl)_AUC _Fresh_T10S_B_ns_sep.08_ AT5G63260   OG0002189 Solyc06g071860.2
```

Fig 1. Example of cross-all traits comparison between Arabidopsis and tomato. First, genes within each trait intersects with its own species reference orthogroups to get the genes included in these orthogroups. Then two gene lists intersect again based on common orthogroups to generate the below result.

Such a comparison generated many gene orthologs results. A huge number of candidate genes generated by cross-all traits comparison are needed to exclude. Therefore, we mainly focused on using genes which directly control or regulate seed germination and dormancy (causal genes) in Arabidopsis. With the help of recent literature review and online databases (Bentsink et al., 2010; Berardini et al., 2015; Carrera-Castaño et al., 2020; Joosen et al., 2012), we obtained 48 causal genes that control or regulate seed germination in Arabidopsis. These genes were used to search their orthologs in tomato for finding candidates that involved in the seed germination.

Manual validation

Once we acquired the candidate genes in tomato, we tracked down these candidates manually in online databases including Uniprot, ENSEMBLPlant, PANTHER (Mi et al., 2018; Morgat et al., 2020; Van Bel et al., 2017; Yates et al., 2020) etc. By comparing different sources and protein function prediction, we determined whether these genes are candidates and their possible function during tomato seed germination.

Permutation test

Although our result showed interesting findings, it is possible that these genes are related because of random matchings on the genome rather than involved in seed germination. To further confirm our results, the randomness of such pairing results was taken into account. First, we calculated the average gene numbers in the QTL region separately for two species. This is to ensure the uniformity of sampling data as well as for the simplicity to run the sampling on the genome. Then equal amount regions on the genomes as the unique QTLs from both species were randomly generated and the genes in these regions were recorded to run the same cross-all traits comparison between species (Fig 2). If Arabidopsis causal genes and its homologs were found within regions comparison, then the number and the name of unique causal genes were documented. Such permutation experiments were executed for 1000 times for testing our previous comparison results.

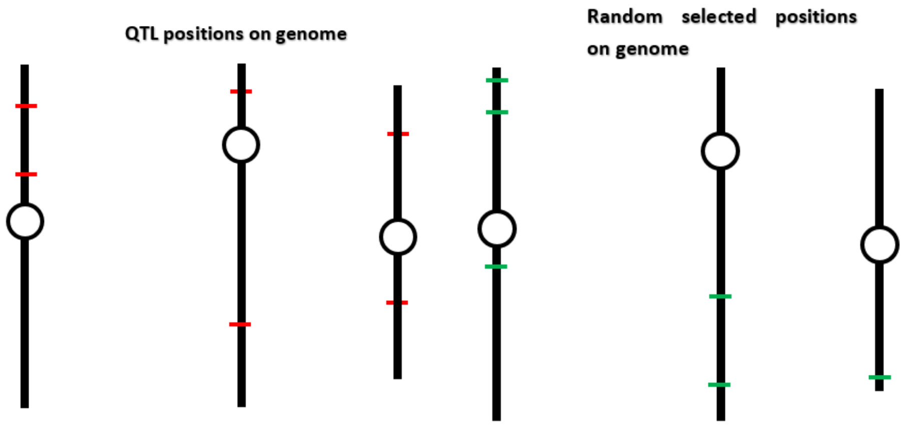


Fig 2. Example of reshuffling regions on genomes for permutation tests. This process is for both Arabidopsis and tomato. The amount and the length of regions are simulated based on real QTL numbers and intervals.

Results

Reference orthogroups

Via OrthoFinder we obtained 11,908 orthogroups where each group included at least one gene from Arabidopsis and from tomato. The comparison yields 27,2279 pairs of genes by cross-all traits comparison. Each line showed genes in the same orthogroups and their related traits. These results contain non-duplicated 2123 tomato genes and 2312 Arabidopsis genes (Table 2). To validate the QTL comparison results and obtain an overview of causal gene candidates in tomato, we first conducted a test using causal genes in Arabidopsis to match the reference orthogroups (Table 3). This test was primarily to validate research feasibility and narrow the number of genes to be investigated.

Sol_traits	Ara_traits	Ara_genes	orthogroups	Sol_genes
Sol_Cold Stress (12 ℃)_AUC	AR_.0_5Mannitol_AUC100_D_ws_okt.10_	AT5G50580	OG0001515	Solyc06g072080.2
Sol_Cold Stress (12 ℃)_AUC	AR_.0_5Mannitol_AUC100_D_ws_okt.10_	AT5G50600	OG0004628	Solyc06g072670.2
Sol_Cold Stress (12 ℃)_AUC	AR_.0_5Mannitol_AUC100_D_ws_okt.10_	AT5G50640	OG0001504	Solyc06g072050.2
Sol_Cold Stress (12 ℃)_AUC	AR_.0_5Mannitol_AUC100_D_ws_okt.10_	AT5G50680	OG0001515	Solyc06g072080.2
Sol_Cold Stress (12 ℃)_AUC	AR_.0_5Mannitol_AUC100_D_ws_okt.10_	AT5G50700	OG0004628	Solyc06g072670.2
Sol_Cold Stress (12 ℃)_AUC	AR_.0_5Mannitol_AUC100_D_ws_okt.10_	AT5G50760	OG0008482	Solyc06g072650.1
Sol_Cold Stress (12 ℃)_AUC	AR_.0_5Mannitol_AUC100_D_ws_okt.10_	AT5G50790	OG0000344	Solyc06g072620.2
Sol_Cold Stress (12 ℃)_AUC	AR_.0_5Mannitol_AUC100_D_ws_okt.10_	AT5G50810	OG0008350	Solyc06g072600.2
Sol_Cold Stress (12 ℃)_AUC	AR_.0_5Mannitol_AUC100_D_ws_okt.10_	AT5G50850	OG0008403	Solyc06g072580.2
Sol_Cold Stress (12 ℃)_AUC	AR_.0_5Mannitol_AUC100_D_ws_okt.10_	AT5G50870	OG0008437	Solyc06g072570.2
Sol_Cold Stress (12 ℃)_AUC	AR_.0_5Mannitol_AUC100_D_ws_okt.10_	AT5G50915	OG0004602	Solyc06g072520.1
Sol_Cold Stress (12 ℃)_AUC	AR_.0_5Mannitol_Gmax_D_ns_okt.10_	AT3G43490	OG0003910	Solyc06g073810.2
Sol_Cold Stress (12 ℃)_AUC	AR_.0_5Mannitol_Gmax_D_ns_okt.10_	AT3G43590	OG0003910	Solyc06g073810.2
Sol_Cold Stress (12 ℃)_AUC	AR_.0_5Mannitol_Gmax_D_ns_okt.10_	AT3G43720	OG0002046	Solyc06g073660.2
Sol_Cold Stress (12 ℃)_AUC	AR_.0_5Mannitol_Gmax_D_ns_okt.10_	AT3G43980	OG0002149	Solyc06g073430.2
Sol_Cold Stress (12 ℃)_AUC	AR_.0_5Mannitol_Gmax_D_ns_okt.10_	AT3G44010	OG0002149	Solyc06g073430.2

Table 2. Example of common gene loci orthogroups between *Solanum lycopersicum* and *Arabidopsis thaliana* derived from seed germination trait QTL. The first two columns represent traits under different treatment. More details concerning conditions and traits were described in these studies (Kazmi et al., 2012; E. A. R. Serin et al., 2017).

We listed 48 genes that are involved in the Arabidopsis germination process based on importance and literature reviews (Berardini et al., 2015; Carrera-Castaño et al., 2020). These genes are major roles in controlling seed germination and dormancy or regulating those who are controlling. 30 of these Arabidopsis genes had corresponding orthologs in tomato (Table 3).

Ara_ID	Ara_genes	Sol_genes
PIL5	AT2G20180	Solyc06g008030.2
PIL5	AT2G20180	Solyc09g063010.2
SOM	AT1G03790	Solyc07g053750.1
SCL3	AT1G50420	Solyc01g008910.2
SCL3	AT1G50420	Solyc12g099900.1
ABA2	AT1G52340	Solyc04g071940.2

ABA2	AT1G52340	Solyc04g071960.2
NCED6	AT3G24220	Solyc05g053530.1
GA20OX1	AT4G25420	Solyc01g093980.2
GA20OX1	AT4G25420	Solyc03g006880.2
GA20OX1	AT4G25420	Solyc06g035530.2
GA20OX1	AT4G25420	Solyc06g050110.1
GA20OX1	AT4G25420	Solyc09g009110.2
GA20OX1	AT4G25420	Solyc09g018410.1
GA20OX1	AT4G25420	Solyc09g018420.1
GA20OX1	AT4G25420	Solyc09g042210.1
GA20OX1	AT4G25420	Solyc10g045690.1
GA20OX1	AT4G25420	Solyc10g045770.1
GA20OX1	AT4G25420	Solyc10g046800.1
GA20OX1	AT4G25420	Solyc10g046820.1
GA20OX1	AT4G25420	Solyc11g013360.1
GA20OX1	AT4G25420	Solyc11g072310.1
ATS2	AT4G30580	Solyc07g005580.2
GA3OX2	AT1G80340	Solyc00g007180.1
GA3OX2	AT1G80340	Solyc01g058250.1
GA3OX2	AT1G80340	Solyc03g119910.2
GA3OX2	AT1G80340	Solyc05g052740.1
GA3OX2	AT1G80340	Solyc06g066820.2
GA3OX1	AT1G15550	Solyc00g007180.1
GA3OX1	AT1G15550	Solyc01g058250.1
GA3OX1	AT1G15550	Solyc03g119910.2
GA3OX1	AT1G15550	Solyc05g052740.1
GA3OX1	AT1G15550	Solyc06g066820.2
GID1B	AT3G63010	Solyc06g008870.2
GID1B	AT3G63010	Solyc09g074270.2
FUS3	AT3G26790	Solyc02g094460.1
CYP707A1	AT4G19230	Solyc08g005610.2
CYP707A1	AT4G19230	Solyc08g075320.2
HAB1	AT1G72770	Solyc03g121880.2
ABI3	AT3G24650	Solyc06g083590.2
ABI3	AT3G24650	Solyc06g083600.1
PER1	AT1G48130	Solyc03g096040.2
ABI5	AT2G36270	Solyc09g009490.2
CHO1	AT5G57390	Solyc07g018290.2
ABA1	AT5G67030	Solyc02g090890.2
ACO1	AT2G19590	Solyc07g026650.2
HDA6	AT5G63110	Solyc03g112410.1
HDA6	AT5G63110	Solyc06g071680.2
KEG1	AT5G13530	Solyc01g096490.2
RGA1	AT2G01570	Solyc10g086370.1
RGA1	AT2G01570	Solyc10g086380.1
RGA1	AT2G01570	Solyc11g011260.1
IMB1	AT2G34900	Solyc09g015660.2

IMB1	AT2G34900	Solyc09g090370.2
HDA9	AT3G44680	Solyc11g067020.1
AIM1	AT4G29010	Solyc07g019670.2
AIM1	AT4G29010	Solyc12g007170.1
HUB1	AT2G44950	Solyc11g013370.1
HUB2	AT1G55250	Solyc01g006030.2
HUB2	AT1G55250	Solyc01g006040.2
EFS	AT1G77300	Solyc04g057880.2
EFS	AT1G77300	Solyc06g059960.2
AL6	AT2G02470	Solyc01g102750.2
AL6	AT2G02470	Solyc01g102760.2
HDA19	AT4G38130	Solyc09g091440.2
BRM	AT2G46020	Solyc01g094800.2
JMJ20	AT5G63080	Solyc03g112600.2
JMJ22	AT5G06550	Solyc10g081630.1

Table 3. common orthogroups of *Solanum lycopersicum* and *Arabidopsis thaliana* based on genes involved in Arabidopsis seed germination.

This primary test suggested that the research goal was feasible and paved the way for searching potential causal genes in tomato at the next step. Hence, we only used these 30 causal genes to match the cross-all traits comparison results. 7 causal genes were found to have orthologs in tomato seed germination QTL (Table 4). SOM, PER1, and GA3OX1 and the corresponding homologs in tomato were found in the current Plant ENSEMBL database (Yates et al., 2020) and showed high WGA (whole genome alignment) coverage scores. While GA20OX1 had no high WGA coverage in the database. GA3OX2 and HDA6 showed no homologs in tomato included by the ENSEMBL database but we found their presence in TAIR (Berardini et al., 2015). Their homologous relations were recorded by these databases. It is noteworthy that only a few publications have been included in the database to support these homologous relations (Fernandez-Pozo et al., 2014; Jun-E Guo et al., 2018; Mi et al., 2018).

Ara_id	Ara_genes	Sol_genes	Function
SOM	AT1G03790	Solyc07g053750.1	Zinc finger transcription factor 50
GA20OX1	AT4G25420	Solyc11g013360.1	Fe2OG dioxygenase domain-containing protein
GA3OX2	AT1G80340	Solyc05g052740.1	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
GA3OX1	AT1G15550	Solyc05g052740.1	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
PER1	AT1G48130	Solyc03g096040.2	1-Cys peroxiredoxin
HDA6	AT5G63110	Solyc06g071680.2	Histone deacetylase
HUB1	AT2G44950	Solyc11g013370.1	E3 ubiquitin protein ligase

Table 4. Candidates of causal genes for tomato seed germination based on common

orthogroups.

We speculated these candidate genes were involved in similar molecular pathways in tomato. In Arabidopsis, SOM downregulates PIL5 which is a light-dependent gene for seed germination (Berardini et al., 2015). PER1 is peroxiredoxin and its expression is limited in seeds and not induced by ABA or drought stress (Berardini et al., 2015). PER1 and its homolog in tomato has a similar function as an antioxidant. However, PER1's tomato homolog A0A3Q7FPK3 only presented in uniformity and t_{50} traits under control conditions. Thus, whether its function is linked with seed germination stress response remains unclear. GA3OX1 and GA3OX2's homologs in tomato are considered as a hypothetical protein that contains a domain in response to the anti-oxidation metabolic pathway (Fernandez-Pozo et al., 2014; Morgat et al., 2020). GA-biosynthesis regulators (like GA3OX1 and GA20OX1) were downregulated in tomato thermo-dormant seeds at elevated temperature (Geshnizjani et al., 2018). However, our results showed both GA3OX1's homolog mainly expressed under oxidative stress and salt stress. No homologs in linked with GA regulators have appeared under heat stress in tomato.

By far, histone acetyltransferases and histone deacetylases are insufficiently researched in tomato seed germination (Jun- E. Guo et al., 2017). Thus, HDA6's homolog presence in tomato seed germination is a very interesting finding. HDA6 is histone deacetylase, a possible candidate for transgene silencing that regulates target genes LEC1/2, FUS3, ABI3 during Arabidopsis seed dormancy and in response to salt stress (Berardini et al., 2015; Carrera-Castaño et al., 2020). The homolog of HDA6 in tomato, A0A3Q7GY95 or SIHDA3(Sato et al., 2012; Zhao et al., 2015) is a histone deacetylase that belongs to RPD3/HDA1 family (Zhao et al., 2015). It has suggested that this homolog can interact with TAG1 gene (Zhao et al., 2015). Depended on limited information (Cigliano et al., 2013; Guo et al., 2017; Sato et al., 2012) it is believed SIHDA3 may be involved in tomato immature green fruit and fruit ripening. Latest research revealed SlHDA3 is also responsible for after ripening senescence (Jun-E Guo et al., 2018). From our results, SIHDA3 appeared generally in 5 germination traits: Gmax, AUC, t_{10}, t_{50}, and MGR under cold stress (12°C), high temperature stress (>35 °C), salt stress, osmotic stress, and oxidative stress conditions in tomato. Up to now, SIHDA3's connection with abiotic stress response suggests that it likely has similar regulation mechanism as HDA6 in Arabidopsis. We speculate that SIHDA3 is also involved in tomato seed germination with the same function in Arabidopsis. Interestingly we did not find LEC1/2 FUS3, and ABI3's homolog presented in tomato seed germination, though all four genes have homologs in tomato (Sato et al., 2012; Yates et al., 2020). In Arabidopsis we noticed SOM is regulated by HDA6-CO-signalling to promote germination (Carrera-Castaño et al., 2020; Jia et al., 2018). Both genes have corresponding homologs in tomato. We suspected that these homologs had similar regulatory relations involved in tomato seed germination. Based on this idea, we searched three more modifiers JMJ20, JMJ22 and BRM which regulate GA3OX1/2 and GA20OX1. These modifiers have homologs in tomato (Table 3). However, no homologs were found within seed germination QTL in tomato. Thus, more evidence is needed to support this hypothesis. DOG1, GID1A, GID1C, and ABI3 directly regulate seed germination in Arabidopsis (Carrera-Castaño et al., 2020). Other important causal genes including ABI5 and LEC1/3 involved in Arabidopsis seed germination showed no corresponding homologous genes in tomato seed germination QTL regions. These genes either like ABI5 which has orthologs in tomato but the orthologs are not present in QTL region or have no common orthologs reported by OrthoFinder. This suggested that tomato may not apply the same ABA pathway to control seed germination. Earlier research showed that the candidate genes which control tomato seed thermo-dormancy were not co-located with the detected QTL (Geshnizjani et al., 2018, 2020). As the modifier genes (i.e. JMJ20) also possess the same situation as ABI5, another explanation is that more QTL regions in tomato were not detected in the previous experiment due to no polymorphism at loci. HUB1 monoubiquitinates histone H2B and regulates seed dormancy (Carrera-Castaño et al., 2020; Liu et al., 2007, p. 1). Its homolog in tomato, SIHUB1(Solyc11g013370), was confirmed by several researches (Cigliano et al., 2013; Zhang et al., 2015; Zhao et al., 2015). These researches indicated that SIHUB1 is involved in response to disease *Botrytis cinereal* (Zhang et al., 2015). As our result showed, SIHUB1 may also be involved in the seed germination process. However, it is difficult to infer which pathway it regulates in seed germination. The existing publications are not sufficient to draw any conclusion on its mechanism yet.

In total, more than 2000 possible homologs in tomato remain to be discovered. For example, tomato's AUC trait under high temperature and salt stress have a common QTL region on chromosome 11 with close marker peak value. Genes from this region have no clear links with causal genes mentioned in Arabidopsis yet. Thus, it is intriguing to investigate novel genes beneath these QTL peaks as well as reversely find more homologs controlling seed germination in Arabidopsis. We also suspect that many QTL in tomato seed germination is controlled more than one gene, as tomato QTL's number is much more than Arabidopsis's and its mechanism is more complicated. Further findings are restrained due to time limitation. More detailed information can be found in the supplementary tables.

QTL regions and causal genes mapping

We collected 952 QTL in Arabidopsis and 118 in tomato. These QTLs are related to seed germination traits under different conditions and time. We removed the duplicated QTL which shared the same regions and resulted in 59 unique QTLs in tomato and 52 in Arabidopsis, respectively. The result also indicated that many traits under the same conditions share common QTL. For example, the QTL region on tomato chromosome 12 near marker peak 44987792, is responsible for the t_{10}, t_{50} and MGR traits under Salt I (-0.3MPa NaCl) stress. By clustering QTL regions that responsible for seed germination traits we can minimize the necessity to find causal genes for repeating QTL regions. These unique QTLs also reduced time needed for the permutation tests.

Conditions	Traits	Chr	Marker peak
Salt II(-0.5MPa NaCl)	Gmax	4	58174884
Salt II(-0.5MPa NaCl)	t_{10}	4	58081284
Salt II(-0.5MPa NaCl)	t_{50}	4	58081284
Salt II(-0.5MPa NaCl)	AUC	4	58174884
Osmotic I(-0.3MPa PEG)	Gmax	4	58174884
Osmotic I(-0.3MPa PEG)	AUC	4	58174884
Control	t_{10}	6	43582592
Control	t_{50}	6	43582592
Control	MGR	6	43582592
Salt I(-0.3MPa NaCl)	t_{10}	6	43582592
Salt I(-0.3MPa NaCl)	t_{50}	6	43582592
Salt I(-0.3MPa NaCl)	MGR	6	43582592
Salt I(-0.3MPa NaCl)	t_{10}	12	44987792
Salt I(-0.3MPa NaCl)	t_{50}	12	44987792
Salt I(-0.3MPa NaCl)	MGR	12	44987792

Table 5. Example of overlapping QTL regions for tomato seed germination.

Permutation test

The 1000 times permutation tests showed that finding 7 causal genes in random regions comparison had low probability, which means our result's p value is <0.001. This directly suggested that our comparison is not out of contingency. Finding these orthologous relations in seed germination traits is not random. Thus, it showed our methods to search causal genes in tomato is feasible. It showed that these 7 gene loci are likely causal genes for seed germination in tomato.

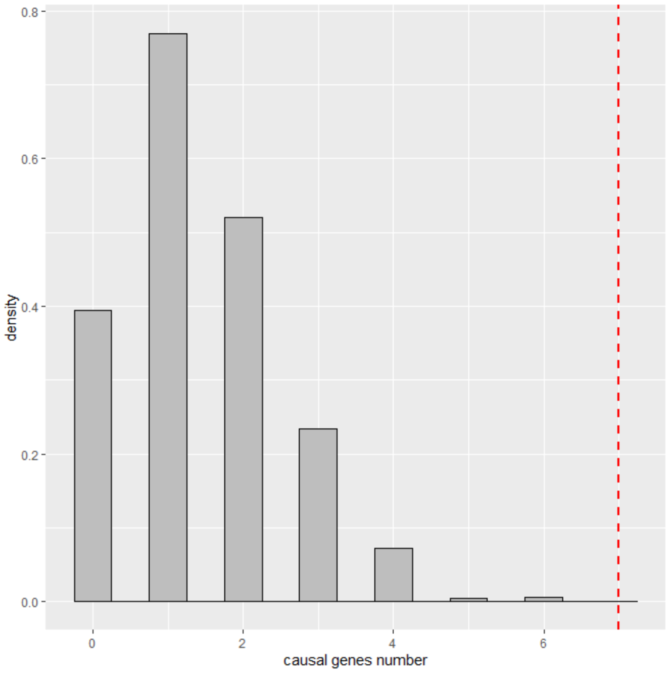


Fig 3. The permutation result of finding number of causal genes in random pairing of regions on genomes between Arabidopsis and tomato. The red dash line represents the 7 unique causal genes detected in this experiment.

Discussion

We found 7 candidate genes for causing tomato seed germination. These genes generally presented in all traits and mainly under stress conditions. Cross-species QTL analysis is regarded as a novel tool for discovering new candidate genes and regulatory pathways based on similar traits. Our experiment proves its feasibility in the first step. The experiment made the most of the existing QTL data. The swift *in-silico* experiment reduced time and lab works and produced valuable and orientational results for further molecular experiments. Our results narrowed down the number of candidates for tomato seed germination. The previous discovery showed none of the genes (FUS3, NCED9, NCED1, GA3ox1 and GA20ox1) which are commonly considered to be involved in the induction of thermo-dormancy in Arabidopsis and lettuce were co-located with these identified QTLs seed germination for heat stress response (Geshnizjani et al., 2018, 2020). Additionally, limited research only provided little information about genes that directly controlling tomato seed germination (Chen et al., 2002; Sato et al., 2012). But by expanding the list of causal genes in Arabidopsis seed germination and using well-constructed orthogroups as references, we managed to find candidates co-located in these QTL regions. These candidate genes were not discovered or thoroughly understood in tomato seed germination prior to our approach. Therefore, our method can contribute to finding more causal genes in tomato efficiently. However, molecular experiments are still needed to confirm our findings. With more testing and confirmation, the method can be promoted to comparing other plant species that manifest similar traits to find more causal genes.

Candidates in tomato are mostly linked with abiotic stress including heat, cold and osmotic stress, suggesting these genes are candidates for stress responses. However, it is still early to conclude that these are candidates for causing seed germination or dormancy. Within our findings, candidate genes of both promoting seed germination and dormancy in Arabidopsis appeared in tomato, thus, regarding these candidates as major causal genes in tomato like DOG1, or ABI5 in Arabidopsis is not confident. While during our online investigation, much conflicting or missing information had to be distinguished. For example, the candidates in tomato were often confirmed on ENSEMBLPlants by whole genome alignment coverage scores, yet they are not included in Uniprot nor their homologous relations are supported by any literature reviews (Morgat et al., 2020; Yates et al., 2020). Another restriction is OrthoFinder itself cannot produce all orthologous relations. On Dicot PLAZA 4.0 (Van Bel et al., 2017) it showed that DOG1 has 3 possible homologs in tomato. But OrthoFinder or other sources (Mi et al., 2018; Yates et al., 2020) showed no such orthologous relations. Another example is LEC1. LEC1 is reported to have 3 homologs in tomato (Sato et al., 2012; Yates et al., 2020). But OrthoFinder failed to detect these homologs. Missing homologs should be manually added to the reference orthogroups to boost more possibilities. Having considered trade-off, we restrained our results within OrthoFinder production and carefully summarized other sources to support our findings. However, this disadvantage does not downgrade the credibility of our results, though it limited the range of possible outcomes. We also considered BLASTp (Madden, 2013) as another tool to yield more possible orthogroups. However, as OrthoFinder has the highest ortholog inference accuracy up to now (Emms & Kelly, 2019). It is only an option to use BLASTp rather than a better confirmation tool. As for the causal genes in Arabidopsis, these 48 listed genes are not complete. Researchers will find more supplementary evidence from databases like TAIR and Dicot PLAZA 4.0(Berardini et al., 2015; Van Bel et al., 2017). We suggested that future research should include more causal genes from Arabidopsis and therefore, give more clear evidence for deducing molecular pathways in tomato.

Although our results are reassured by permutation, it is still possible that these candidates are not the real causal genes in tomato. Apart from listed causal genes, these QTL regions still contain many orthologous genes left to be discovered. One possible way to filter these genes is to check whether they are expressed during seed germination. AraQTL provides a convenient way to check every locus expression in different traits (Nijveen et al., 2017). Once a similar database for tomato is finished, it will be convenient to check these genes expression levels before initiate experiments. In the future, with the help of these databases, filtering out redundant and non-expression loci can speed up causal genes findings.

As for the length of QTL interval, it is a rough estimation of actual QTL size. We must emphasize that such a 2Mb interval is not a true representative of every QTL regions, even though some intervals are mentioned in the data. Another aspect is that QTL of different traits in different species varies its length a lot. One research pointed out that QTL interval in rice can vary from 200kb-3Mb and the number of underlying genes is limited to less than 450 (Bargsten et al., 2014). During our experiment, we found tomato QTL regions held 230 genes less than Arabidopsis on average. In this experiment we used a 2Mb interval only for simplicity and uniformity. Similar experiments in the future should consider more thoroughly for different QTL. Besides these biological factors, the quality of QTL mapping also holds importance. The Arabidopsis QTL data (Serin et al., 2017) has good quality to guarantee our experiments' reliability.

Besides seed germination traits, seed quality is also part of seed performance. Seed size and seed weight influence tomato seed quality the most and has a clear correlation (Geshnizjani et al., 2020). Our findings have not connected tomato seed quality traits with Arabidopsis seed quality traits due to time limitations. But we argued that such a method could be applied directly to seed quality traits.

Lastly, this experiment executes on existing data to discover the other less known plant. We managed to find an important and well-studied trait in Arabidopsis, seed germination to give a first insight of causal genes in tomato. We believe such method is promising because of its efficiency in terms of time and statistically reliable results.

Reference

Argyris, J., Dahal, P., Hayashi, E., Still, D. W., & Bradford, K. J. (2008). Genetic variation for lettuce seed thermoinhibition is associated with temperature-sensitive expression of abscisic acid, gibberellin, and ethylene biosynthesis, metabolism, and response genes. *Plant Physiology*, *148*(2), 926–947.

Bargsten, J. W., Nap, J.-P., Sanchez-Perez, G. F., & van Dijk, A. D. (2014). Prioritization of candidate genes in QTL regions based on associations between traits and biological processes. *BMC Plant Biology*, *14*(1), 330.

Bentsink, L., Hanson, J., Hanhart, C. J., Blankestijn-de Vries, H., Coltrane, C., Keizer, P., El-Lithy, M., Alonso-Blanco, C., de Andrés, M. T., 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*(9), 4264–4269. https://doi.org/10.1073/pnas.1000410107

Berardini, T. Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., & Huala, E. (2015). The arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. *Genesis*, *53*(8), 474–485. https://doi.org/10.1002/dvg.22877

Carrera-Castaño, G., Calleja-Cabrera, J., Pernas, M., Gómez, L., & Oñate-Sánchez, L. (2020). An Updated Overview on the Regulation of Seed Germination. *Plants*, *9*(6), 703.

Chen, F., Nonogaki, H., & Bradford, K. J. (2002). A gibberellin-regulated xyloglucan endotransglycosylase gene is expressed in the endosperm cap during tomato seed germination. *Journal of Experimental Botany*, *53*(367), 215–223.

Churchill, G. A., & Doerge, R. W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics*, *138*(3), 963–971.

Cigliano, R. A., Sanseverino, W., Cremona, G., Ercolano, M. R., Conicella, C., & Consiglio, F. M. (2013). Genome-wide analysis of histone modifiers in tomato: Gaining an insight into their developmental roles. *BMC Genomics*, *14*(1), 57.

Cui, J., Luan, Y., Jiang, N., Bao, H., & Meng, J. (2017). Comparative transcriptome analysis between resistant and susceptible tomato allows the identification of lnc RNA 16397 conferring resistance to Phytophthora infestans by co-expressing glutaredoxin. *The Plant Journal*, *89*(3), 577–589.

Emms, D. M., & Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology*, *20*(1), 238. https://doi.org/10.1186/s13059-019-1832-y

Falconer, D. S. (1996). *Introduction to quantitative genetics*. Pearson Education India.

Fernandez-Pozo, N., Menda, N., Edwards, J. D., Saha, S., Tecle, I. Y., Strickler, S. R., Bombarely, A., Fisher-York, T., Pujar, A., Foerster, H., Yan, A., & Mueller, L. A. (2014). The Sol Genomics Network (SGN)—From genotype to phenotype to breeding. *Nucleic Acids Research*, *43*(D1), D1036–D1041. https://doi.org/10.1093/nar/gku1195

Geshnizjani, N., Ghaderi-Far, F., Willems, L. A., Hilhorst, H. W., & Ligterink, W. (2018). Characterization of and genetic variation for tomato seed thermo-inhibition and thermo-dormancy. *BMC Plant Biology*, *18*(1), 229.

Geshnizjani, N., Sarikhani Khorami, S., Willems, L. A., Snoek, B. L., Hilhorst, H. W., & Ligterink, W. (2019). The interaction between genotype and maternal nutritional environments affects tomato seed and seedling quality. *Journal of*

Experimental Botany, *70*(10), 2905–2918.

Geshnizjani, N., Snoek, B. L., Willems, L. A. J., Rienstra, J. A., Nijveen, H., Hilhorst, H. W. M., & Ligterink, W. (2020). Detection of QTLs for genotype × environment interactions in tomato seeds and seedlings. *Plant Cell and Environment*, *43*(8), 1973–1988. https://doi.org/10.1111/pce.13788

Guo, J.-E., Hu, Z., Guo, X., Zhang, L., Yu, X., Zhou, S., & Chen, G. (2017). Molecular Characterization of Nine Tissue-Specific or Stress-Responsive Genes of Histone Deacetylase in Tomato (Solanum lycopersicum). *Journal of Plant Growth Regulation*, *36*(3), 566–577. https://doi.org/10.1007/s00344-016-9660-8

Hormozdiari, F., Kichaev, G., Yang, W.-Y., Pasaniuc, B., & Eskin, E. (2015). Identification of causal genes for complex traits. *Bioinformatics (Oxford, England)*, *31*(12), i206–i213. PubMed. https://doi.org/10.1093/bioinformatics/btv240

Huo, H., & Bradford, K. J. (2015). Molecular and hormonal regulation of thermoinhibition of seed germination. In *Advances in plant dormancy* (pp. 3–33). Springer.

Jia, Y., Li, R., Yang, W., Chen, Z., & Hu, X. (2018). Carbon monoxide signal regulates light-initiated seed germination by suppressing SOM expression. *Plant Science*, *272*, 88–98.

Joosen, R. V. L., Arends, D., Willems, L. A. J., Ligterink, W., Jansen, R. C., & Hilhorst, H. W. M. (2012). Visualizing the Genetic Landscape of Arabidopsis Seed Performance. *Plant Physiology*, *158*(2), 570–589. https://doi.org/10.1104/pp.111.186676

Kazmi, R. H., Khan, N., Willems, L. A., Van Heusden, A. W., Ligterink, W., & Hilhorst, H. W. (2012). Complex genetics controls natural variation among seed quality phenotypes in a recombinant inbred population of an interspecific cross between Solanum lycopersicum× Solanum pimpinellifolium. *Plant, Cell & Environment*, *35*(5), 929–951.

Kearsey, M. J. (1998). The principles of QTL analysis (a minimal mathematics approach). *Journal of Experimental Botany*, *49*(327), 1619–1623.

Khan, N., Kazmi, R. H., Willems, L. A., Van Heusden, A. W., Ligterink, W., & Hilhorst, H. W. (2012). Exploring the natural variation for seedling traits and their link with seed dimensions in tomato. *PLoS One*, *7*(8), e43991.

Lander, E. S., & Botstein, D. (1989). Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, *121*(1), 185–199.

Lin, F., Fan, J., & Rhee, S. Y. (2019). QTG-Finder: A machine-learning based algorithm to prioritize causal genes of quantitative trait loci in Arabidopsis and rice. *G3: Genes, Genomes, Genetics*, *9*(10), 3129–3138.

Liu, Y., Koornneef, M., & Soppe, W. J. J. (2007). The Absence of Histone H2B Monoubiquitination in the Arabidopsis hub1 Mutant Reveals a Role for Chromatin Remodeling in Seed Dormancy. *The Plant Cell*, *19*(2), 433. https://doi.org/10.1105/tpc.106.049221

Mackay, T. F. (2001). The genetic architecture of quantitative traits. *Annual Review of Genetics*, *35*(1), 303–339.

Madden, T. (2013). The BLAST sequence analysis tool. In *The NCBI Handbook [Internet]. 2nd edition*. National Center for Biotechnology Information (US).

Meadows, D. H., Meadows, D. L., & Randers, J. (1992). *Beyond the limits: Confronting global collapse, envisioning a sustainable future*. Post Mills, Vt.: Chelsea Green Pub. Co.,.

Mi, H., Muruganujan, A., Ebert, D., Huang, X., & Thomas, P. D. (2018). PANTHER version 14: More genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Research*, *47*(D1), D419–D426. https://doi.org/10.1093/nar/gky1038

Miles, C., & Wayne, M. (2008). Quantitative trait locus (QTL) analysis. *Nature Education 1 (1)*, *208*.

Morgat, A., Lombardot, T., Coudert, E., Axelsen, K., Neto, T. B., Gehant, S., Bansal, P., Bolleman, J., Gasteiger, E., de Castro, E., Baratin, D., Pozzato, M., Xenarios, I., Poux, S., Redaschi, N., & Bridge, A. (2020). Enzyme annotation in UniProtKB using Rhea. *Bioinformatics*, *36*(6), 1896–1901. https://doi.org/10.1093/bioinformatics/btz817

Nijveen, H., Ligterink, W., Keurentjes, J. J., Loudet, O., Long, J., Sterken, M. G., Prins, P., Hilhorst, H. W., De Ridder, D., & Kammenga, J. E. (2017). Ara QTL–workbench and archive for systems genetics in Arabidopsis thaliana. *The Plant Journal*, *89*(6), 1225–1235.

Nunes-Nesi, A., Alseekh, S., de Oliveira Silva, F. M., Omranian, N., Lichtenstein, G., Mirnezhad, M., González, R. R. R., y Garcia, J. S., Conte, M., & Leiss, K. A. (2019). Identification and characterization of metabolite quantitative trait loci in tomato leaves and comparison with those reported for fruits and seeds. *Metabolomics*, *15*(4), 46.

R Core Team. (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rueedi, R., Ledda, M., Nicholls, A. W., Salek, R. M., Marques-Vidal, P., Morya, E., Sameshima, K., Montoliu, I., Da Silva, L., & Collino, S. (2014). Genome-wide association study of metabolic traits reveals novel gene-metabolite-disease links. *PLoS Genetics*, *10*(2).

Sato, S., Tabata, S., Hirakawa, H., Asamizu, E., Shirasawa, K., Isobe, S., Kaneko, T., Nakamura, Y., Shibata, D., Aoki, K., Egholm, M., Knight, J., Bogden, R., Li, C., Shuang, Y., Xu, X., Pan, S., Cheng, S., Liu, X., … Universitat Pompeu, F. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*, *485*(7400), 635–641. https://doi.org/10.1038/nature11119

Serin, E. A. R., Willems, L. A. J., Nijveen, H., Snoek, L. B., Hilhorst, H. W. M., & Ligterink, W. (2017). *Genetic variation in the effect of seed maturation environment on seed performance*. International society for seed science: Triennial Conference 2017.

Serin, Elise A. R., Snoek, L. B., Nijveen, H., Willems, L. A. J., Jiménez-Gómez, J. M., Hilhorst, H. W. M., & Ligterink, W. (2017). Construction of a High-Density Genetic Map from RNA-Seq Data for an Arabidopsis Bay-0 × Shahdara RIL Population. *Frontiers in Genetics*, *8*(201). https://doi.org/10.3389/fgene.2017.00201

Van Bel, M., Diels, T., Vancaester, E., Kreft, L., Botzki, A., Van de Peer, Y., Coppens, F., & Vandepoele, K. (2017). PLAZA 4.0: An integrative resource for functional, evolutionary and comparative plant genomics. *Nucleic Acids Research*, *46*(D1), D1190–D1196. https://doi.org/10.1093/nar/gkx1002

Voorrips, R. E., Verkerke, W., Finkers, R., Jongerius, R., & Kanne, J. (2000). Inheritance of taste components in tomato. *Acta Physiologiae Plantarum*, *22*(3), 259–261.

Weirauch, M. T., Yang, A., Albu, M., Cote, A. G., Montenegro-Montero, A., Drewe, P., Najafabadi, H. S., Lambert, S. A., Mann, I., & Cook, K. (2014). Determination and inference of eukaryotic transcription factor sequence specificity. *Cell*, *158*(6), 1431–1443.

Yao, C.-W., Hsu, B.-D., & Chen, B.-S. (2011). Constructing gene regulatory networks for long term photosynthetic light acclimation in Arabidopsis thaliana. *BMC Bioinformatics*, *12*(1), 335. https://doi.org/10.1186/1471-2105-12-335

Yates, A. D., Achuthan, P., Akanni, W., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M. R., Armean, I. M., Azov, A. G., Bennett, R., Bhai, J., Billis, K., Boddu, S., Marugan, J. C., Cummins, C., Davidson, C., Dodiya, K., Fatima, R., Gall, A., … Flicek, P. (2020, January 8). *Ensembl 2020*. Nucleic Acids Res.

Zhang, Y., Li, D., Zhang, H., Hong, Y., Huang, L., Liu, S., Li, X., Ouyang, Z., & Song, F. (2015). Tomato histone H2B monoubiquitination enzymes SlHUB1 and SlHUB2 contribute to disease resistance against Botrytis cinerea through modulating the balance between SA-and JA/ET-mediated signaling pathways. *BMC Plant Biology*, *15*(1), 1–20.

Zhao, L., Lu, J., Zhang, J., Wu, P.-Y., Yang, S., & Wu, K. (2015). Identification and characterization of histone deacetylases in tomato (Solanum lycopersicum). *Frontiers in Plant Science*, *5*, 760.