Quantitative Trait Locus and Network Analysis for Light Response in Tomato Seed Germination

Margi Hartanto
Registration number: 910819307060
Major MSc Thesis Plant Physiology (PPH-80436)

Supervisor:
Nafiseh Geshnizjani MSc

Examiner:
dr. ir. Wilco Ligterink

Laboratory of Plant Physiology
Wageningen University and Research
January 2018

ACKNOWLEDGEMENTS

This thesis was not a single-person task. Many people helped me and contributed so much to this work and I would like to express my gratitude towards them.

I got a lot of valuable experience from this thesis. So first, I want to thank Wilco Ligterink for the opportunity given so that I am allowed to carry out my major thesis in Laboratory of Plant Physiology. His guidance and support through many discussions had helped me understand the difficult parts of this work.

Second, I want to thank my daily supervisor, Nafiseh Geshnizjani, as she always had time for discussion and tirelessly answering my questions during the thesis period. Despite her business, I am really thankful for her patient and willingness to teach me some trivial works.

Third, I want to thank Leo Willems for his technical support and remarkable effort from the beginning until the end of this thesis. I realize that I am very clumsy in wet lab work but he always be patient to help me and I am very grateful for that.

I also want to thank the people who work and study in Laboratory of Plant Physiology, my friends, and my parents for their invaluable support during this thesis period.

Finally, special thanks to LPDP (Indonesia Endowment Fund for Education) for the financial support, thus I can continue my education at Wageningen University and Research.

ABSTRACT

Tomato (*Solanum lycopersicum* cv. Moneymaker) seed germination is regulated by light. It is known that this environmental factor is perceived by a photoreceptor protein called phytochromes. However, the gene downstream to this signal perception has not been discovered. Therefore, in this study, we carried out a quantitative trait loci analysis using a recombinant inbred line population of *Solanum lycopersicum* x *Solanum pimpinellifolium*. This analysis was then followed by gene co-expression network construction to identify the candidate gene for light germination. Here we show that the parental line under study has different seed germination behavior as the response to light, particularly in terms of germination onset (t10) and rate (t50). Furthermore, QTL analysis revealed twenty-nine QTLs specific for light germination trait, where most of the beneficial alleles were coming from S. pimpinellifolium. Subsequent network analysis for Gmax and AUC trait also lead to the identification of four candidate genes: Solyc04g051720, Solyc05g014090, Solyc05g042180, and Solyc06g069710. These candidate genes were then validated for their involvement in light germination by expression analysis using qPCR and to database search for their Arabidopsis orthologous. As the result, Solyc06g069710 was chosen as the most promising candidate gene for light germination. In summary, we show that QTL mapping followed by gene co-expression network analysis can be utilized to dissect the genetic control of quantitative trait in plants.

Keywords: genetical-genomics, network analysis, seed germination, QTL analysis, tomato

ABREVIATION:

AUC = area under the curve; the integration of fitted germination curve between t = 0 and a user-defined end-point
cFr = continuous far-red light
CL = continuous light, regardless of the spectrum
cR = continuous red light
eQTL = expression quantitative trait loci
Fr = Far-red light
Frp = Far-red light pulse
Gmax = maximum germination capacity of a seed lot in percentage
MM = *Solanum lycopersicum* cultivar Moneymaker
PIM = *Solanum pimpinellifolium*
QTL = quantitative trait locus
R = Red light
Rp = Red light pulse
t10 = seed germination onset; time required for 10% of viable seeds to germinate
t50 = seed germination rate; time required for 50% of viable seeds to germinate
u8416 = uniformity; the time interval between 84% and 25% of viable seeds to germinate

Table of content

1. INTRODUCTION

1.1. Seed germination quality

A good production of agronomic and horticultural crops largely depends on the quality of seed. These complex traits include a wide range of parameter, starting from nutrition content to germination characteristic (Wimalasekera, 2015). In the context of germination characteristics, high-quality seeds could be recognized by several parameters such as high germination percentage, uniformity, and speed. These attributes, in the end, contribute to a high yield and quality of the crop. Therefore, breeding for seed quality is important in modern agricultural crops. However, breeding for a quantitative trait such as seed quality is rather difficult because of the major influence of environmental factors, not to mention the complex genetic control.

Numerous environmental effects play an essential role in determining the quality of seed germination. This is not without reasons as the basis of environmental regulation in seed germination is to determine the most appropriate time for seedlings establishment in a particular condition (Finch-Savage & Leubner-Metzger, 2006). Therefore, beneficial seed germination traits in plants have been selected in nature as an adaptive response to the adverse environmental condition. The same mechanism also occurred in modern cultivated crops, even though the selection is mostly unintentional and only for traits that are apparent for crop propagation (Martínez-Andújar et al. 2012). As the consequence, there might be a possibility that a certain trait in the modern crop is accidentally selected against without our knowledge.

1.2. The influence of light in seed germination with particular emphasis in tomato

Light is an important factor that regulates seed germination in many plant species. However, the requirement of light for germination is different among plant species. Many species, such as *Arabidopsis* (Shropshire *et al*., 1961) and *Lactuca sativa* cv. Grand Rapids (Borthwick et al. 1952; Borthwick et al. 1954), required light, particularly in the spectrum of red light, to initiate seed germination. On the other hand, the presence of light is not an absolute requirement of seed germination. This could be observed in tomato which seed germinates readily in the dark (Mancinelli et al. 1966). However, these examples might be an oversimplification as the response of seed germination to light depend on the quality and the quantity of the light itself.

In term of quality, a different spectrum of light has a different effect on seed germination. This was shown by Borthwick, Hendricks and, Parker (1952) in lettuce cv. Grand Rapid. They figured out that light in the spectrum of far-red light might suppress the germination of lettuce seed while red light spectrum promotes it. Later work in Arabidopsis also suggested the same thing (Shropshire et al., 1961b). These results led to the finding of light absorbing protein pigment, namely phytochrome. This pigment has two type of form: Pr, a red light absorbing form, and Pfr, a far-red light absorbing form. Phytochrome presence in the Pr form at the time it is synthesized, hence the presence of Pr is abundant in the newly developed seed. This pigment is photoreversible, indicating that exposure to red light can convert this pigment to Pfr form and subsequent far-red light exposure may convert it back to Pr form (Taiz & Zeiger 2010).

In term of quantity, the response of seed germination to light could be classified into three phase depending on the quantity of photon or fluence of R that the seed received (Casal et al. 1998) (Bewley et al. 2013). The first is very-low fluence response (VLFR) where the low ratio of Pfr/Ptotal is sufficient to induce this response. This response is not photoreversible as the exposure of Frp can also induce this response due to overlapping action spectra with red light. A higher ratio of Pfr/Ptotal induces the second response, which is called low fluence response (LFR). In contrast to VLFR, LFR is photoreversible, hence the promoting effect of R can be reversed by Fr. The last response is high irradiance response (HIR) which is, in comparison to the other two responses, depend on the light irradiance and not obey the law of reciprocity. This response is induced by continuous irradiation to Fr in seed germination.

Due to its favorable seed size and the availability of phytochrome mutants, tomato is frequently used as a model plant species in the study of light response in seed germination (Lercari & Lipucci di Paola 1991; van Tuinen et al. 1995; Shichijo et al. 2001; Appenroth et al. 2006). The earliest study on the relationship between light and tomato seed germination was carried out by Mancinelli, Borthwick, and Hendricks in 1966. They discovered that single short far-red pulse (FRp) applied briefly after the seed is rehydrated can inhibit the germination. This inhibitory effect of Fr on seed germination can be reversed by immediate exposure to red pulse (Rp). Furthermore, if the FRp exposure is delayed hours after imbibition, the suppressive effect of Fr on seed germination is decreased. Hourly exposure of Frp or continuous exposure of Fr (cFR), even at the time when Frp is no longer effective, however, bring back the inhibitory effect of this light spectrum (Shichijo et al. 2001). The same suppressive effect is also shown by hourly exposure of Rp or continues exposure to red light (cR) (Appenroth et al. 2006). This difference indicates the presence of LFR and HIR in tomato seed germination.

The LFR and HIR in tomato seed germination have distinct photoreceptor for light perception. In LFR, the inhibitory effect of Fr and the promoting effect of R are regulated by *PHYB2* (Appenroth et al. 2006), and possibly by *PHYB1* (Bertram et al. 2004). Seed germination in the dark is also under the regulation of *PHYB2* and *PHYB1* since the mutant of these gene has low germination percentage in the absence of light (Bertram et al. 2004). On the other hand, suppression of seed germination as the response of high irradiance of R or Fr are mediated by *PHYA* (Shichijo et al. 2001; Appenroth et al. 2006).

Even though the signal perception mechanism has been discovered, the downstream signaling of R or Fr in tomato seed germination is not fully elucidated. The inhibitory effect of Fr might act in the micropylar region of imbibed tomato seed where galactomannan-hydrolyzing enzyme activity is suppressed after Frp (Nomaguchi et al. 1995). The light stimulus might also influence the expression of several genes. For instance, *GIGANTEA*, a gene known to be involved in light signaling, is upregulated as the response of Frc (Auge et al. 2009). However, the expression profile of GIGANTEA needs to be confirmed as another work also showed that Rc could upregulate the expression of this gene (Rutitzky et al. 2009).

1.3. The utilization of genetical-genomics in studying quantitative traits

As previously mentioned, quantitative traits such as seed quality have complex genetic control because they are regulated by multiple genes. As the consequence, the discovery of the underlying genetic and molecular mechanism of such traits is more difficult to perform compared to a single gene trait. An alternative method to study quantitative traits could be provided by quantitative trait locus analysis. This is performed by testing the association of phenotypic and genotypic value of plants in a segregating population, leading to the identification of regions within the genome which contain genes associated with specific quantitative trait (Collard et al. 2005). This kind of analysis has been proven successfully in the genetic study of quantitative trait and even lead to gene identification by further reverse genetic approach (as reviewed by Morgante and Salamini, 2003; Paran and Zamir, 2003).

More sophisticated analysis could also be performed by treating genome-wide expression data as phenotypic traits. This analysis results in the mapping of eQTLs: polymorphic loci that cause a differences in gene expression. These eQTLs could be classified into two groups based on the location of the causal polymorphism. The first one is local *cis*-eQTL where difference in gene expression is caused by allelic variation of the gene itself. If the cause of polymorphism is located not inside the gene under study, then it is called *trans* eQTL. This type of eQTL could also be grouped into local-*trans* eQTL, where the

polymorphism is within the eQTL confidence interval, and distant-*trans* eQTL, where the polymorphism is located outside the eQTL confidence interval (Joosen et al. 2009). Among these group of eQTL, *cis* eQTLs is of great importance in candidate gene identification. This is due to the fact that they are co-located with phenotypic QTL and differentially expressed (Druka et al. 2010). Another approach could also be performed by combining whole eQTL data with co-expression analysis and functional classification to construct a gene co-expression network (Jiménez-Gómez et al. 2010).

1.4. Research scope and research objective

The light response to tomato seed germination has been an interesting topic of research in recent decades. Previous works show that multiple photoreceptor pigments involved in light perception, leading to promotion or inhibition of seed germination. However, our knowledge in the downstream signaling pathway is still limited. In order to elucidate the molecular mechanism of light-mediated seed germination in tomato, we are going to exploit the potency of natural variation found in inbred lines derived from a crossing between a cultivated tomato species, *Solanum lycopersicum* cv. Moneymaker, and its wild relative, *Solanum pimpinellifolium*. Previous work by Kazmi et al. (2012) using the same inbred lines results in numerous QTLs for 42 seed quality traits. Using the same approach, we are going to study the genetic control of seed quality under different light condition. In addition to QTL analysis, we will also make use of eQTL data to see which gene is differentially expressed under the influence of the phenotypic QTL. Eventually, this approach leads to a research question for this study: what gene(s) is involved in tomato seed germination under a different light condition?

2. MATERIALS AND METHODS

2.1. Plant material

Seeds were obtained from an *F8* RIL population developed from a cross between *Solanum lycopersicum* cv. Moneymaker (referred later as MM)—a well-known tomato cultivar—with *Solanum pimpinellifolium* (referred later as PIM), an inbreed wild-species of tomato. The genotype for 865 single nucleotide polymorphisms (SNPs) of all lines was obtained from an *F7* population.

2.2. Germination assay

A randomized complete block design was used to determine the mean value of seed quality of 101 RILSs. Each genotype was replicated 4 times, which consist of two biological replicates and two technical replicates. The germination assay was carried out in germination trays (21 x 15 cm; DBP Plastics NV, Antwerpen, Belgium, http://www.dbp.be) layered with two blue blotter paper (Anchor paper company, size 5.6 X 8 inches). Each germination trays contained three replicates; each consist of 40—50 seeds. The trays were hydrated with 50 ml of demineralized water. No more than 15 trays then were piled up. To prevent unequal water evaporation, two empty trays, layered with two white blotter paper and watered by 75 ml demineralized water, were put on the top and the bottom position of the germination tray pile. The whole tray then was wrapped in a transparent plastic bag and placed in a cold room at 4° C for 3 days for stratification. Next, all piles of the tray were incubated in an incubator at 25° C under continuous light condition. Lastly, the scoring of seed germination was performed for 6 days with the frequency of twice a day for the first three days and once a day for the next three days.

2.3. Statistical analysis

The parameters of seed quality measured in this study are Gmax, t10, t50, u8416, and AUC. These parameters were estimated using curve-fitter module of Germinator package (Joosen et al. 2010). Mean of the four replicates were calculated and used for QTL analysis. Particularly for Gmax, the means of all replicates were normalized using probit transformation prior further analysis.

2.4. Heritability measurement

For genetic analysis, the value of broad-sense heritability was calculated to know the proportion of genetic variance. The total phenotypic variance of each trait was divided into genotypic variance (V_G) and error

variance (V_E) using one-way analysis of variance (IBM SPSS Statistics 23). These variance components then were used to calculate broad-sense heritability as $H^2 = V_G/(V_A+V_G)$.

2.5. QTL analysis for seed quality in light germination condition

QTL analysis was carried out based on the established marker linkage map of RIL population consisting 865 SNP markers. This analysis was performed using computer program MapQTL version 6 (Van Ooijen 2009). First, putative QTLs were identified by performing simple interval mapping. Next, a marker close to each detected QTL peak was selected as a cofactor to carry out multiple-QTL mapping analysis. The resulting QTL peaks and the nearest marker to each of it were then determined.

2.6. Gene co-expression network construction

There is a high possibility that the polymorphism of the candidate gene causes expression difference of this and another gene downstream to the candidate gene. Based on this assumption, we then compiled a set of gene having shared eQTL peak in the location of marker closest to the phenotypic QTL. TomQTL (http://altschul.bioinformatics.nl/SolQTL/) was used to carry out this analysis. Since the database is still in the development at the time of this study, only expression profile of dried seed between the two species under study (Geshnijzani et al. 2017, publication in preparation) is available. However, we assume that the difference in gene expression in dried seed also influences the seed germination particularly under different light condition.

The expression profile of genes having shared eQTL in the QTL peak was provided from TomExpress (Zouine et al. 2017; tomexpress.toulouse.inra.fr) by first limiting the output only from seed and seedling of both mutant and wild-type tomato under biotic and abiotic stresses. I used Spearman Rank Correlation to determine the correlation coefficient of every pair of genes in the dataset. A modified R script written by Rashid Kazmi was used to perform such analysis. The resulting list containing correlation value between pairs of genes was then used to construct gene co-expression network using Cytoscape 3.5.1 (Shannon et al. 2003). Prior network construction, a threshold for correlation value was determined in a way to eliminate non-significant edges without removing too many nodes. This is usually ranging from 0.8 to 0.95 and different from one set of gene to another.

2.7. RNA Isolation and cDNA synthesis

An experiment was designed to validate the expression of the candidate gene in germinating *MM* cv. Moneymaker and *PIM* under dark and continuous light condition at 6 and 12 hours after imbibition. The

seeds used in this experiment were also obtained from the different maternal environment to see whether there is an interaction between environment and light condition of for the gene expression. Therefore, there were 2 genotype x 2 light conditions x 2 time points x 2 maternal environment x 3 replicates = 48 unit of experiments. For each experiment, around 30 seeds were used for RNA isolation. The seeds were grinded with tissue lyser until it turned into a fine white powder. The sample powder was used for total RNA extraction using phenol extraction method as performed previously by Dekkers *et al.* (2012). The isolated RNA was then stained using EtBr and assessed for the quality using electrophoresis on 1% agarose gel. The sample that has degraded RNA is subsequently discarded. Furthermore, the total RNA concentration was measured as well using Trinean Xpose spectrophotometer (BIOKÉ). The sample that has a low concentration of RNA or has a ratio of OD260/OD280 outside the range of 1.8 to 2.2 was discarded and the RNA isolation was repeated for such sample. Lastly, cDNA was synthesized using iScript^{TM} cDNA synthesis kit (Bio-Rad).

2.8. Real-time quantitative polymerase chain reaction (RT-qPCR)

The next step was the quantification of the candidate genes mRNA by RT-qPCR using SYBR Green as the fluorescence reporter. As the reference gene, SGN-U567355 were included based on their stable expression in tomato seed (Dekkers et al. 2012). As much as 12.5 µl cDNA template was mixed with 0.5 µl of 10 µM reverse primer, 0.5 µl of 10 µM forward primer, and 12.5 µl iQ Sybr Green Supermix (Bio-Rad). The reaction consists of 3 cycles. The first cycle took place at 95**º** C for 3 minutes. Next, 40 cycles were performed which consist of 95**º** C for 15 seconds and 60**º** C for a minute. At this time, cycle threshold (Ct) values were collected. The reaction continues in 60**º** C that is increased as much as 0.5**º** C every 5 seconds until it reaches 95**º** C to gather melting curve data. The Ct values were then obtained from CFX Manager Software (Bio-Rad). These values were used to calculate the relative gene expression using Livak Method with Ct of dried seed in HP as the calibrator (Livak & Schmittgen 2001).

2.9. Primer design

The primers for the candidate gene were designed based on the sequence of the candidate genes retrieved from SOL Genomics Network and generated using CLC Main Workbench 7. Either the forward or the reverse primer were spanned the exon-intron border to prevent the amplification of the genomic DNA. We also set the melting temperature ranging from 55**º** C to 62º C and GC content to 40% to 60%. The primer efficiency was assessed by determining its Ct value on 4 different dilution factors. Only primer sets with 90% to 100% efficiency are used in this study.

The primer is going to be assessed for its quality by amplifying them under the above-mentioned qPCR protocol and visualize the product by electrophoresis on 1% agarose gel to check whether the primer complementary sequence is unique and not forming primer-dimer. More test will be carried on by running another PCR cycles to obtain the melting curve for each primer set.

3. RESULT

3.1. Trait evaluation

The effect of light in seed germination is varied among plant species. To explore this natural variation as the response to light in two tomato species, seeds of *Solanum lycopersicum* cv. Moneymaker, *Solanum pimpinellifolium*, and their F8 RILs population were germinated in complete dark and continuous light condition. Subsequently, they were measured for Gmax, t10, t50, U8416, and AUC. These results were then summarized and shown in Figure 1 and Figure 2.

Two parental lines tested showed a relatively high germination percentage as indicated by Gmax values with a narrow standard error. Even though lower than the parental lines, the average Gmax of RIL population was also considered as high yet the standard error was a bit large, indicating the presence of Gmax variation among lines in the population. The high variation was not only observed in LN but also in HP maternal environment.

Under different lighting condition, however, the behavior of seed germination onset—indicated by t10—and germination rate—indicated by t50—of two parents were not the same. For MM in light treatment, the average of t10 and t50 was significantly higher ($P<0.05$) compared to dark treatment. In the contrary, seeds of PIM germinate faster if treated with light, implied by a significantly lower average of t10 and t50 ($P<0.05$). Although, in HP, the difference is only shown for t10 but not t50. Despite this, the result in general obviously demonstrated the variation of response in seed germination onset and speed as the effect of light stimuli.

As for the RILs population, the average value of t10 and t50 was located in between of both parents. Furthermore, no significant difference between lighting conditions was observed on these parameters of RIL population. In terms of germination uniformity (indicated by u8416), the mean was lower in dark compared to dark treatment, yet the differences once again were not significant. Eventually, the different response toward the light in seed germination onset and rate apparently affect the AUC of both parents as this value was significantly higher in dark treatment for MM and lower in PIM ($P<0.05$).

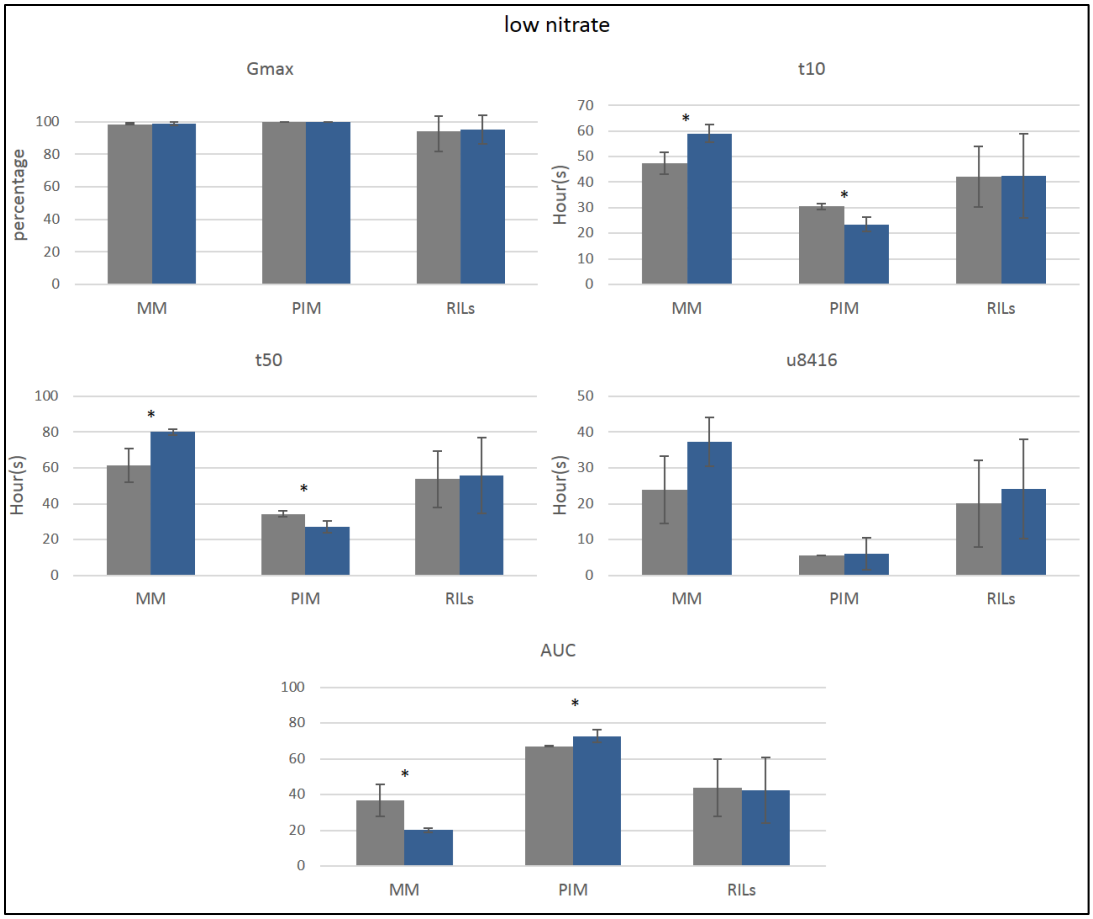


Figure 1. mean germination trait (±se) of *Solanum lycopersicum* cv. Moneymaker, *Solanum pimpinellifolium,* and F8 RIL population derived from the crossing of both species. The data were taken under dark (grey bar) and light (blue bar) conditions in LN maternal environments. Mean values with an asterisk (*) shows significant differences between light and dark treatment according to t-test ($p<0.05$).

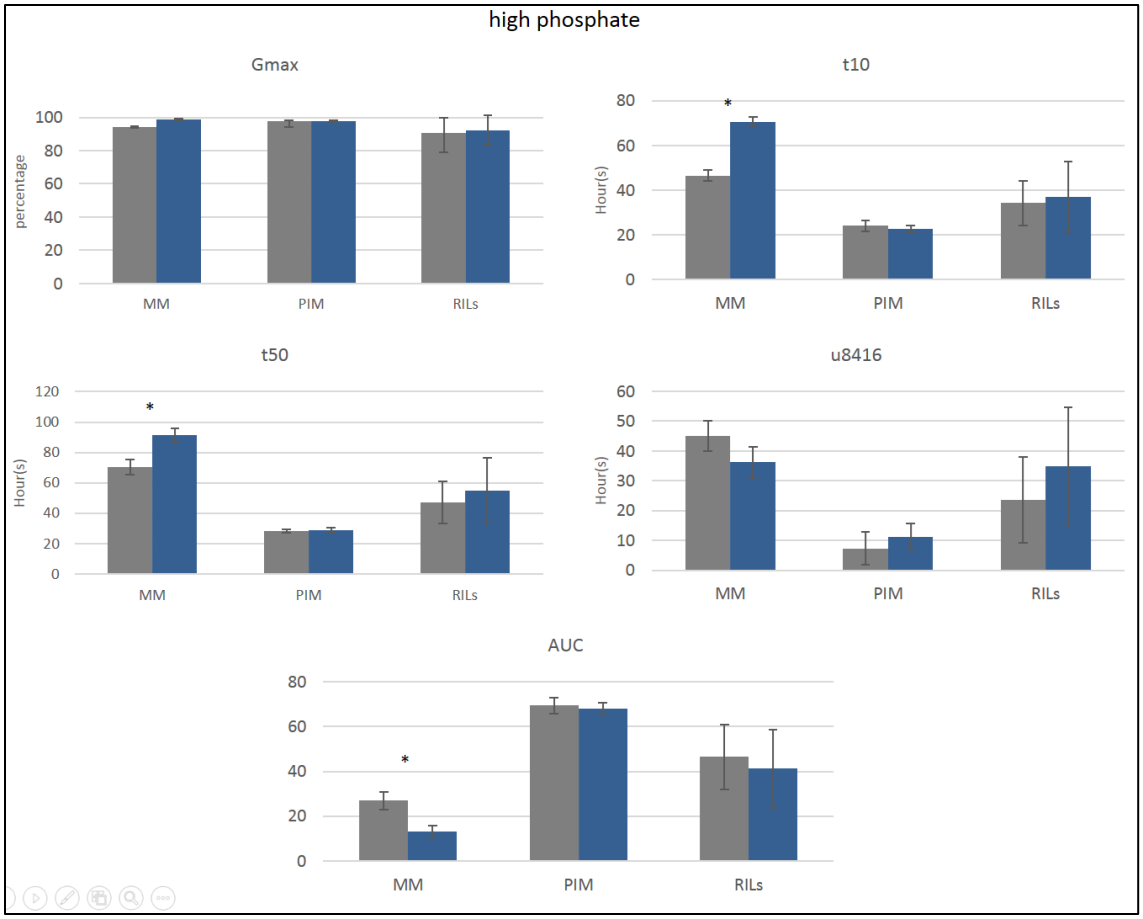


Figure 2. mean germination trait (±se) of *Solanum lycopersicum* cv. Moneymaker, *Solanum pimpinellifolium,* and F8 RIL population derived from the crossing of both species under dark (grey bar) and light (blue bar) conditions in HP maternal environments. Mean values with different letter show significant differences between light and dark treatment according to t-test ($p<0.05$). Mean values with an asterisk (*) shows significant differences between light and dark treatment according to t-test ($p<0.05$).

3.2. Frequency distribution of RILs population

The phenotype histograms for genotypes under study visualize the segregation of RILs population relative to both parental lines (Figure 3). As expected from quantitative traits, the approximately normal distribution was observed on the histogram of all studied trait, including Gmax where no significant differences between the mean of both parental lines were shown. Moreover, in most traits, there were RILs having a transgressive phenotype in the positive or negative direction relative to both parents' phenotypes. These transgressive phenotypes were more observed in the traits where both parents had a more similar phenotype, such as in Gmax under both maternal environments, t10 under LN, and u8416 under HP.

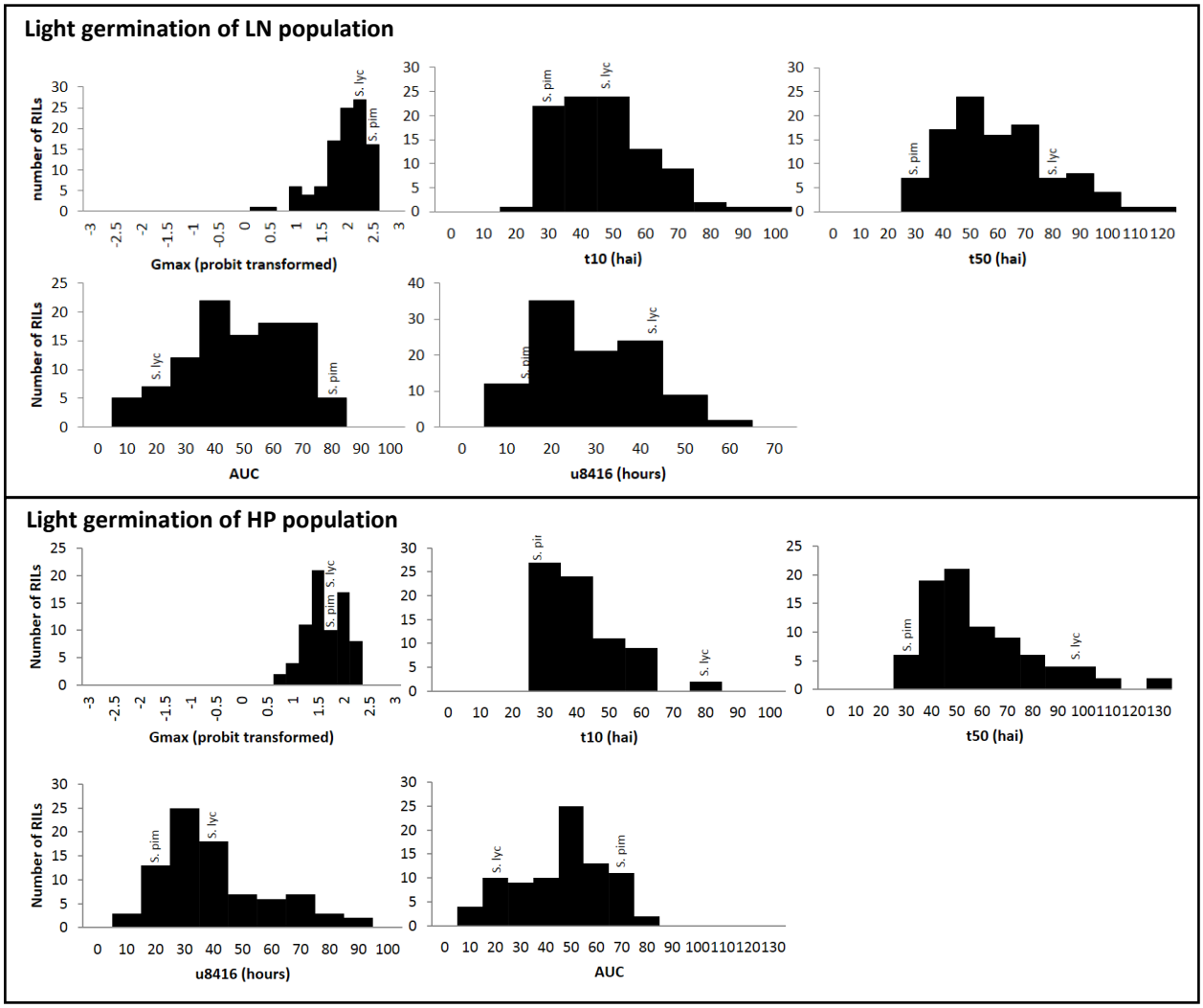


Figure 3. Frequency distribution of all germination characteristic traits in the *Solanum lycopersicum* x *Solanum pimpinellifolium* RIL population under light condition. The average parental value is indicated by S. lyc and S. pim for MM and PIM, respectively. AUC: area under the curve.

3.3. QTL analysis for seed germination traits

Quantitative trait locus analysis was carried out to find genetic loci associated with seed germination traits as the response to light treatment. The first step for this was the identification of putative QTLs using LOD threshold of 2.0 through simple interval mapping. Next, a set of marker located near to the putative QTLs was used as cofactors in MQM mapping. At this step, I only chose a marker that increases the LOD in order to discover more segregating QTLs. Finally, a restricted MQM mapping was carried out to acquire the final one-support interval and LOD score of discovered QTLs.

A total of 62 QTLs were found to be associated with dark or light germination traits under two different maternal environment (Table 1). Even though most QTLs were collocating for a certain trait in two lighting condition, as much as 29 QTLs were found to be associated specifically to light seed germination trait. On average, these QTLs explained approximately one-third of the phenotypic variance. The highest was for Gmax in light germination in HP maternal environment where 56.1% of the variance was explained by four QTLs, which is considered one of the largest in this study. This is found to be quite surprising since the heritability for this trait is only 37.61%, which is the lowest among all traits under study. As the matter of fact, low heritability may indicate the minimum influence of the genetic factors for the phenotypic variation of Gmax in HP; at least it is the case in this study. Low heritability was also obtained for Gmax of dark germination in the same environment, which is only 65.51%, yet two QTL are identified for this trait. On the other hand, the heritability for Gmax in LN environment is relatively high, around 87.6% for dark germination and 83.3% for light germination, but only 1 QTL for dark and 2 QTLs for light germination.

Besides Gmax, the percentage of explained variance and heritability were also determined for the other traits. Traits corresponding to seed germination onset and seed germination rate (t10 & t50) have a relatively high heritability, ranging from 82% to 90% in HP and 84% to 94% in LN. A considerable number of QTLs were also identified for these traits. Six QTLs were associated for t10 and t50 of dark germination in HP and five of them were collocating with t10 and t50 of light germination in the same environment. However, the other one non-collocating QTL is still located on the same chromosome (chromosome 3) and separated by around 70 cM of genetic linkage. On the other hand, nineteen QTLs were identified for germination speed in LN. Several of these, however, were low effect QTLs, explaining only 5—6% of phenotypic variance. Some QTLs were also collocated with the same trait in HP environment, for instance, QTLs for t10 of dark germination was collocated with QTLs for t10 of dark germination in HP environment.

In LN maternal environment, three QTLs were associated with each u8416 trait in dark and light germination condition.

QTLs were also found for u8416, a trait that represents uniformity. The genomic region in chromosome 5 within an interval around 20—35 cM solely explained 10—11% of the variation in u8416 for both dark and light germination in HP maternal environment. However, this region is not associated with u8416 in LN environment, which suggest the specificity of interaction between the gene underlying this QTL and the environmental factors in regulating germination uniformity. Furthermore, three QTLs were associated with the u8416 trait in dark and light germination and only one is collocating with each other, which is located in chromosome 11 (supp. int. 21—32 cM).

As for AUC, seven QTLs were found to be associated with this trait in HP; five for dark germination and two for light germination. The two QTLs for AUC of light germination were also present in the dark germination so there was no specific QTLs for this trait. Meanwhile, under the influence of LN maternal environment, three QTLs were identified for AUC of dark germination and four for AUC of light germination. Interestingly, one of the QTLs for AUC, which is located in the interval around 43—58 cM of chromosome 11 was also associated for Gmax, t10, and t50 of light germination in LN. This may indicate the pleiotropic effect of gene underlying the QTL.

In general, each chromosome has at least one QTL except for chromosome 2 where no single QTL was found. Moreover, there are several regions in the genetic map that has an association for more than one traits. For instance, a region in chromosome 7 (supp. int: ≈23–36 cM) and 11 (supp. int: ≈20–30 cM) that are associated with more than seven traits under the influence of different maternal environment. These QTLs have a negative effect, suggesting that the beneficial allele is coming from PIM. Beside this, there are also QTLs that were associated with seed germination traits but only identified in specific light condition and maternal environment. An example for this is QTL in chromosome 12 that is associated with t50, u8416, and AUC in light condition and LN maternal environment. Another example was provided by QTL in chromosome 8 that was associated with t10, t50, u8416, and AUC in dark condition under LN maternal environment. These examples depict the complex and diverse interaction between the genetic and environmental factors in the regulation of seed germination.

In contrary to previous findings, seven unique QTLs were found, which associated only with one specific germination trait in a particular maternal environment. An example is a QTL in chromosome 4 that has support interval spanning from 55.1 cM to 59.4 cM. This QTL is exclusively associated with Gmax in light

condition and HP maternal environment. This QTL had an LOD of 5.51 accounted for 24.9% explained variance, which was the highest among QTLs discovered in this analysis. The beneficial allele apparently coming from PIM by examining the negative effect of the QTL. The closest marker to the peak of this QTL was located on 51.6 bp, almost collocating with *SEP2* that was suspected to be involved in cellular response to high light intensity. Beside this, another trait-specific QTLs can be found in chromosome 3 (supp. int: 22.4–31.7 cM) for t10 of dark germination in HP, chromosome 3 (supp. int: 121.9–123.9 cM) for t10 of light germination in LN, chromosome 4 (supp. int: 55.1–59.4 cM) for t10 of light germination in LN, chromosome 5 (supp. int: 66.9–83.4 cM) for Gmax of light germination in HP, chromosome 9 (supp. int: 52.4–60.9 cM) for t50 of dark germination in LN, and chromosome 12 (supp. int: 51.7–58.1 cM) for t10 of light germination in LN.

Table 1. Chromosomal location of QTLs associated with germination characteristic traits of *Solanum lycopersicum* x *Solanum pimpinellifolium* RIL population under dark and light conditions with different maternal environments

Maternal environment	Trait	Light condition	Chr[1]	Marker peak[2]	Support interval[3] (cM)	LOD[4]	Explained variance[5] (%)	Total explained variance[6] (%)	Effects[7]	Heritability[8] (%)
HP	Gmax	dark	5	59856707	62.7-70.8	2.6	9.9	19.4	-0.18	65.51
			6	32090367	18.3-34.5	2.55	9.5		0.186	
	Gmax	light	4*	51678315	55.1-59.4	5.51	24.9	56.1	-0.2	37.61
			5	60619277	66.9-83.4	3.01	10.4		-0.13	
			6	31371391	9.9-24.5	2.72	11.2		0.13	
			7*	28076245	26.4-36.4	2.35	9.6		-0.12	
	t10	dark	3	48283991	22.4-31.7	4.01	12.2	43.4	3.54	84.13
			6	35748330	54.3-59.4	2.36	8		-2.9	
			6	43580814	87.7-102.1	3.27	10.9		-3.35	
			11	48283991	21.9-32.7	4.01	12.3		3.69	
	t10	light	7*	28076245	20.4-32.4	2.14	10.1	22.4	4.8	82.66
			11	49292479	14.3-35.1	2.55	12.3		4.98	
	t50	dark	3	58231771	101.5-118.3	2.46	7.7	38.2	3.86	90.62
			6	35748330	54.4-58.7	2.2	7.7		-3.6	
			6	43580814	87.7-102.1	2.82	9.8		-4.01	
			11	48283991	24.4-34.13	3.99	13		4.62	
	t50	light	7*	28076245	22.4-35.4	2.66	12.6	23.7	7.52	87.23
			11	47543161	20.85-34.1	2.36	11.1		6.63	
	u8416	dark	5	266350	0-8	2.59	11.2	11.2	-10.2	68.66
	u8416	light	5	266350	0-8	2.43	10.6	10.6	-10.47	70.14
	AUC	dark	3	seq-rs5561	111.3-119.3	3.66	10.2	46.6	-4.6	91.58
			5	59856707	62.7-66.9	3.57	9.9		-4.54	
			6	32202956	21.1-45.1	2.65	7.2		3.8	
			7	28076245	27.4-37.4	3.3	9.1		-4.23	
			11	48596795	24.4-33.7	3.67	10.2		-4.54	
	AUC	light	7	28076245	23.4-35.4	2.08	8.9	18.2	-4.86	87.17
			11	47543161	14.3-27.8	2.17	9.3		-4.95	

LN	Gmax	dark	7	28076245	26.4-35.4	2.23	9.8	9.8	-0.18	87.60
	Gmax	light	5*	6814373	51.4-63.2	2.54	9.3	23	-0.14	83.30
			11*	48586795	19.2-30.1	3.64	13.7		-0.17	
	t10	dark	3	61881530	121.9-123.9	4.54	10.3	53.9	-5.83	87.49
			5	7031168	56.0-59.7	3.31	7.3		-4.76	
			6	43580814	92.0-104.3	2.22	5.2		3.8	
			8	58825288	84.3-89.4	6.75	16.2		-6.55	
			11	47543161	21.3-25.4	6.31	14.9		-6.66	
	t10	light	4*	5460003	42.0-55.7	2.87	8.3	44.7	-4.42	84.72
			5*	416826	0-10.0	2.09	5.9		-3.78	
			11	48283991	21.9-27.8	6.06	18.7		6.87	
			12*	47269280	51.7-58.1	4.05	11.8		5.56	
	t50	dark	1	74314733	91.9-119	2.15	4.9	51.5	3.9	91.33
			6	43580814	94.9-100.8	4.99	12.4		-5.47	
			8	58825288	83.3-88.4	5.81	14.7		6.06	
			9	58088185	52.4-60.9	2.57	6		-3.9	
			11	48283991	23.4-30.1	5.37	13.5		5.84	
	t50	light	4*	4936160	41.0-51.5	3	8.3	50.9	-5.96	94.85
			4*	55204967	67.3-73.6	3.05	8.4		-5.93	
			5*	6711222	51.4-65.9	2.24	5.8		4.94	
			11	48283991	21.9-31.7	6.02	17		8.63	
			12*	6238531	45.9-54.8	4.23	11.4		6.97	
	u8416	dark	6	43580814	92.8-102.1	3.74	11.5	32.7	-3.86	75.28
			8	58825288	83.3-89.4	2.56	7.6		3.24	
			11	48283991	21.9-30.7	4.38	13.6		4.3	
	u8416	light	7*	28076245	24.4-39.8	2.57	8.4	35.4	3.33	92.87
			11	49292479	25.4-32.7	5.35	18.7		5.02	
			12*	6238531	43.9-53.8	2.55	8.3		3.39	
	AUC	dark	5	59786600	53.3-93.1	2.16	7.7	26.5	3.1	92.44
			7	28076245	23.4-39.8	2.23	8		3.15	
			8	58825288	76.4-89.4	2.99	10.8		3.93	
	AUC	light	1*	76672459	99.7-119.8	2.26	5.1	42.6	-4.19	96.86
			6*	43761285	92.8-102.1	3.03	8.3		5.23	
			11*	48283991	23.4-30.7	7.03	21.2		-8.41	
			12*	6238531	43.9-58.1	2.93	8		-5.23	

Gmax (%): maximum germination percentage; t10 and t50 (h^{-1}): time required for germination of respectively 10% and 50% of the total imbibed seeds; u8416: time interval between the germination of 16% to 84% of the total imbibed seeds; AUC: area under the curve, integration of fitted curve between the start of imbibition to the end point; QTL: quantitative trait locus; RIL: recombinant inbreed lines, LOD: logarithm-of-odds. [1]Chromosome number; [2]The name and physical location of the nearest marker to the QTL peak; [3]1-LOD support interval of QTL; [4] [10]log of likelihood ratio that the association occurs by linkage than by chance; [5]Percentage of variation explained by individual QTL;[6]Percentage of variation explained by genetic factors estimated by MapQTL; [7]Effect of QTL calculated as µA - µB, where A and B are RILs carrying *MM* and *PIM* alleles at the QTL position, respectively. µA and µB were estimated by MapQTL. Effects are given in percentage (Gmax) and h^{-1} (t10, t50, MGR, U7525-1). [8] Broad-sense heritability estimate for each trait, estimated as the proportion of phenotypic variance explained by genotype in a one-way analysis of variance model; calculated as $H^2 = V_G/V_P + V_E$

3.4. Gene co-expression network construction and candidate gene determination

Finding the candidate gene from each phenotypic QTLs would be very important for further researches in functional studies. This, nevertheless, is not a simple task since the gene is hidden somewhere in a wide genomic region. Therefore, I utilized a genetical genomics approach to find the candidate genes by combining the QTL and gene expression data from the publicly available database. From this, the gene co-expression networks were constructed, opening the possibilities to discover the most possible candidate gene from seed germination trait QTLs. In this study, I performed the network analysis for QTLs of Gmax and AUC that were found specifically in the light condition.

The first step in the network construction was to find genes which expression are affected by a polymorphism occurred near to the peak of phenotypic QTL. The "Shared peaks" window in TomQTL was used for this purpose. In general, hundreds of gene IDs were detected for each phenotypic QTL, however, not all of these genes could be proceeded due to lack of gene name, duplication (two gene name with the same gene ID), and multiple gene names corresponded to the same gene IDs. This has become a bottleneck factor in this study as the candidate gene might not be discovered due to this limitation. Nevertheless, the available genes were then proceeded and summarized and the expression data for each of these genes were sought.

Table 2. The number of genes used for the construction of gene co-expression network.

Maternal environment[1]	Trait	Chromosome	Peak locus (mb) [2]	Genes sharing eQTL[3]	Genes processed[4]	Threshold for coexpression analysis[5]	Genes having strong correlation[6]	Trans acting gene[7]			Cis acting gene[10]	
								Total	diff. exp. [8]	located in other chr. [9]	Total	diff. exp. [8]
LN	Gmax	4	51,678,315	516	158	0.8	73	44	41	21	29	29
LN	Gmax	7	28,076,245	303	137	0.8	102	76	21	2	26	26
HP	Gmax	5	6,814,373	434	213	0.95	61	33	7	2	27	27
HP	Gmax	11	48,586,795	247	166	0.85	93	82	37	9	10	10
HP	AUC	1	76,672,459	305	143	0.8	101	93	39	6	7	7
HP	AUC	6	43,761,285	136	127	0.85	75	65	27	12	10	10
HP	AUC	11	48,283,771	247	90	0.85	89	88	44	8	1	1
HP	AUC	12	6,238,531	281	166	0.85	81	56	18	3	25	25

[1]The environment of maternal plant, [2] This is based on the physical location of marker closest to the QTL peak,[3] Number of genes that have an eQTL peak on the phenotypic QTL peak,[4] Genes available to be processed for co-expression analysis [5] R^2 threshold for co-expression data for network construction [6] Number of gene that have correlation to another gene with R^2 above the threshold [7] Number of trans regulated gene predicted by its physical location outside the support interval of phenotypic QTL [8] Number of gene that have different expression across genotype and maternal environment corresponded to the phenotypic QTL [9] [10] Number of cis acting gene predicted by its physical location in the QTL support interval

In principle, genes involved in a certain biological pathway usually have an identical expression pattern under various genotype background and environmental condition. Hence, TomExpress database was used to obtain an insight about co-expressed genes in the previous dataset. This database allows us to mine for public RNA-Seq data of multiple genes at once. There is also an option to filter only to show the expression only from the specific developmental stages, genotype, or stress conditions. For this study, the expression array was collected from seed and seedling stage of mutants and wild-type of PIM and MM under several different abiotic conditions (Table 3).

Table 3. Source of RNA-seq data used for the expression array

Species	Strain	Cultivar	Organ Type	Organ	Tissue	Stage	Treatment	Duration	Pathogen
S pimp	Wild Type	LA.1589	Reproductive	Seed	Embryo	4.DPA	-	-	-
S pimp	Wild Type	LA.1589	Reproductive	Seed	Endosperm	4.DPA	-	-	-
S pimp	Wild Type	LA.1589	Reproductive	Seed	Seed.Coat	4.DPA	-	-	-
S pimp	Wild Type	LA.1589	Reproductive	Seed	Funiculus	4.DPA	-	-	-
S lyco	Wild Type	SUN1642	Reproductive	Seed	Whole	7.dpa	-	-	-
S lyco	sun_SA4_overexpressed_mutant	SUN1642	Reproductive	Seed	Whole	7.dpa	-	-	-
S lyco	Wild Type	SUN1642	Reproductive	Seed	Whole	10.dpa	-	-	-
S lyco	sun_SA4_overexpressed_mutant	SUN1642	Reproductive	Seed	Whole	10.dpa	-	-	-
S lyco	Wild Type	Micro.Tom	Reproductive	Seed	Whole	IMG_10DPA	-	-	-

S lyco	Wild Type	Micro.Tom	Reproductive	Seed	Whole	MG_35DPA	-	-	-
S lyco	Wild Type	Micro.Tom	Reproductive	Seed	Whole	Breaker_38DPA	-	-	-
S lyco	Wild Type	Micro.Tom	Reproductive	Seed	Whole	Orange_41dpa	-	-	-
S lyco	Wild Type	Micro.Tom	Reproductive	Seed	Whole	Red_44DPA	-	-	-
S lyco	Wild Type	M82	Vegetative	Seedling	Whole	10.dpg	sun	-	-
S lyco	Wild Type	M82	Vegetative	Seedling	Whole	10.dpg	shade	-	-
S lyco	Wild Type	M82	Vegetative	Seedling	Whole	10.dpg	shade	-	-
S lyco	Wild Type	M82	Vegetative	Seedling	Leaf	2.young.leaves	PAC	3d	-
S lyco	Wild Type	M82	Vegetative	Seedling	Leaf	2.young.leaves	PAC_GA	3d_30min	-
S lyco	proDeltaGRAS.mutant	M82	Vegetative	Seedling	Leaf	2.young.leaves	PAC	3d	-
S lyco	proDeltaGRAS.mutant	M82	Vegetative	Seedling	Leaf	2.young.leaves	PAC_GA	3d_30min	-

Taken from TomExpress

At this point, I already had the expression array of all genes, each consist of 20 RNA-Seq data. These data were then used to construct gene co-expression network. First, the similarity between each possible pair of genes was measured by Spearman's Rank Correlation. This type of correlation test was chosen due to its sensitivity to outliers and suitability to the non-linear relationship. After the similarity scores were calculated, the threshold for the R^2 values was determined. The threshold value could be different among each dataset, depending on the number of genes in it. For instance, the dataset for QTL on chromosome 5 had 213 genes in it, thus a stringent selection was applied, or else the resulting network would be hard to be interpreted due to the presence of too many nodes and insignificant edges. The threshold value for each dataset was compiled in Table 2.

After the co-expression data were acquired, the network analysis was performed and visualized using Cytoscape 3.5.1 (Shannon et al. 2003). In order to make it easier to be interpreted, two attributes was additionally given to each gene in the network. The first one was based on the genes physical location relative to the phenotypic QTL support interval to identify a gene with cis-eQTL in the region as an indication of the candidate gene. In the network, genes located in the QTL support interval were represented by bordered nodes.

The second gene attribute was given upon its expression pattern corresponding to the trait and the QTL under study. The source of gene expression was derived from RNA-seq data of dried seed (Nafiseh et. al. unpublished). An example of this attribution will be given using a network for Gmax QTL under HP maternal environment in chromosome 4, where the positive allele are originated from PIM. We can assume that, if the candidate gene for this QTL were the positive regulator for seed germination, then it should show higher expression in both PIM compared to MM and in HP relative to LN maternal environment. The same pattern is also applied for the repressor of Gmax, but the expression should be lower. Furthermore, the downstream gene to the regulation of the candidate gene should have the same expression pattern due to their involvement in the same pathway. These co-expressed genes were then

identified and prioritized in the network to indicate that these genes were transcribed altogether correspond to the trait under study.

From eight networks constructed, only three networks resulted in promising candidate genes for light germination (Figure 4—6). The first one was from QTL located on chromosome 4 for Gmax in HP maternal environment (Figure 4). The clustering algorithm for genes sharing an eQTL in this chromosomal location resulted in three large and several small networks. One of these networks were filled with several nodes representing gene with eQTL in the phenotypic QTL location and surrounding the nodes of Solyc04g051720. This gene is collocated with phenotypic QTL and among the nodes with high degree in the network, suggesting that this gene might be a good candidate for the trait under study.

To figure out the relation between Solyc04g051720 and the co-expressed gene in germination under light condition, the information related to this gene and its orthologues was retrieved from the publicly available database (Table 4). Based on PANTHER (http://www.pantherdb.org/), Solyc04g051720 is an orthologue to AT5G64820 and AT1G16850 genes in Arabidopsis, which are described as a transmembrane protein. This gene is co-expressed with 10 other genes in the network. Among these genes, there is Solyc08g078190, which has high homology with AT4G17490. Interestingly, this gene encodes for an ethylene transcription factor as a response to high intensity of light.

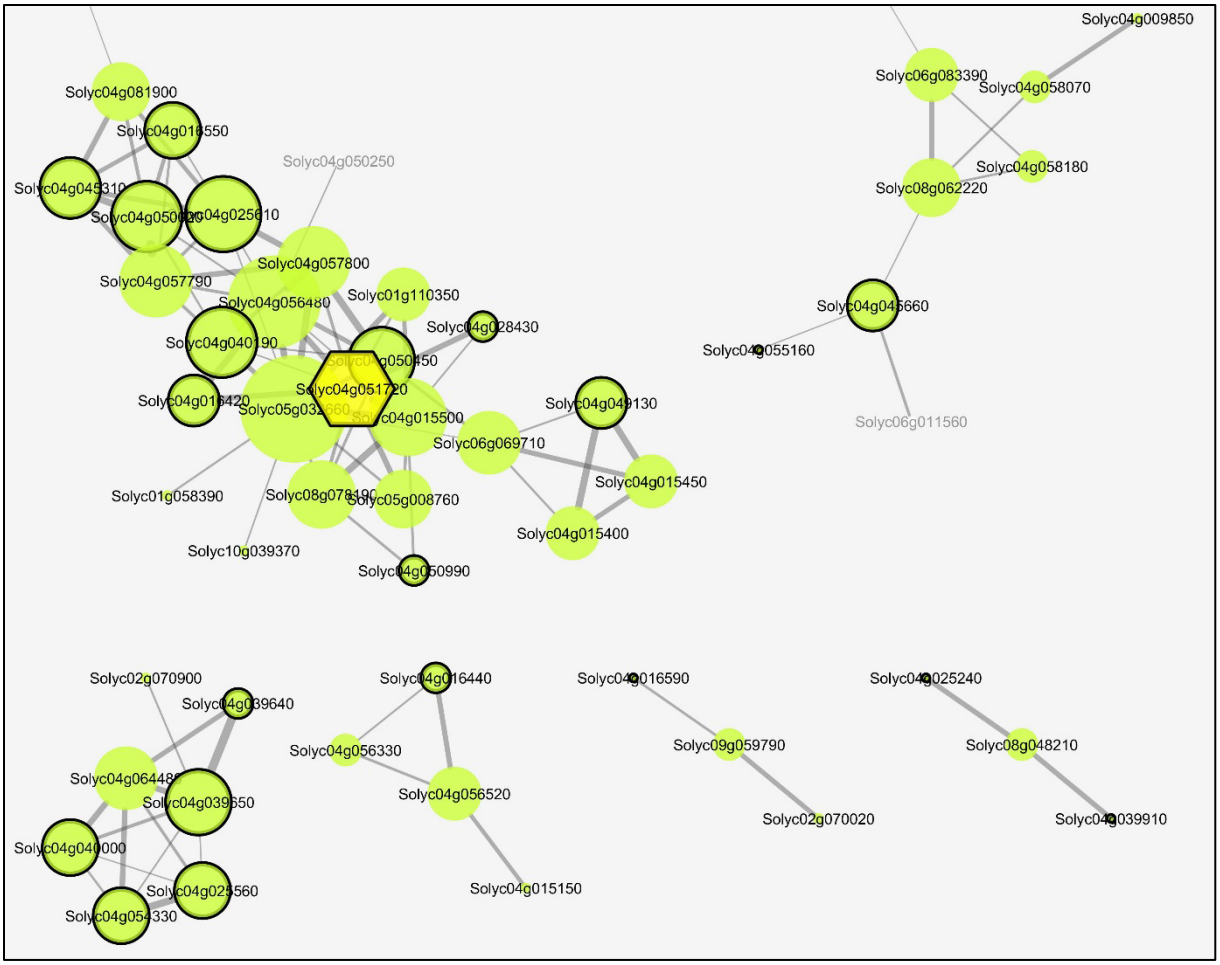


Figure 4. Network analysis for 73 genes located in the QTL in chromosome 4 for Gmax in HP maternal environment (peak locus: 51,678,315 mb). Genes are represented by nodes. Nodes with borders are genes located in the support interval of the QTL on chromosome 4. Larger nodes have more degree of connection compared to smaller nodes. Moreover, nodes with green color represent genes that have higher expression (or lower if it is an inhibitor) in dried PIM seed under HP maternal environment compared to the other genotype and maternal environment. These genes are connected to each other by edges. Thick edges indicate strong correlation and the other way around for thin edges. The candidate genes for this QTL is Solyc04g051720, which is represented by a yellow hexagon-shaped node that is connected to the other ten nodes. Note that several insignificant nodes and edges are not shown in the figure to maintain the image resolution.

Table 4. Description of Solyc04g051720 as a candidate gene for Gmax under HP in chromosome 4 and its orthologues in Arabidopsis

candidate gene

Gene	cis-eQTL	connection degree	description	orthologue in Arabidopsis	description
Solyc04g051720	yes	10	**SGN**: Genomic DNA chromosome 5 P1 clone MXK3 (AHRD V1 ***- Q9LV95_ARATH)	AT1G16850	**TAIR**: transmembrane protein
				AT5G64820	**TAIR**: transmembrane protein

co-expressed gene

Gene	cis-eQTL	correlation coefficient	description	orthologue in Arabidopsis	description
Solyc05g032660	no	-0.94	**SGN**: Dehydrogenase/ reductase 3 (AHRD V1 *--* Q3INM6_NATPD); contains	AT5G04900	**TAIR**: Encodes a chlorophyll b reductase involved in the degradation of chlorophyll b and LHCII (light-harvesting complex II)

			Interpro domain(s) IPR002347 Glucose/ribitol dehydrogenase; **Pfam**: short chain dehydrogenase, Enoyl-(Acyl carrier protein) reductase, KR domain		
Solyc04g015500	no	0.93	SGN: B3 domain-containing protein At3g18960 (AHRD V1 *-*- Y3896_ARATH); contains Interpro domain(s) IPR003340 Transcriptional factor B3; **Pfam**: B3 DNA binding domain	AT3G18960	**TAIR**: AP2/B3-like transcriptional factor family protein
				At4g01580	**TAIR**: AP2/B3-like transcriptional factor family protein
				AT1G49475	**TAIR**: AP2/B3-like transcriptional factor family protein
Solyc04g028430	yes	0.88	**SGN**: Unknown Protein (AHRD V1)	-	-
Solyc05g008760	no	-0.88	**SGN**: Amino acid permease (AHRD V1 ***- B6SYL4_MAIZE); contains Interpro domain(s) IPR002293 Amino acid/polyamine transporter I; **Pfam**: Amino acid permease	AT2G01170	**TAIR**: Encodes a bidirectional amino acid transporter that can transport ala, arg, glu and lys, GABA but not pro with both export and import activity. Its expression is localized in the vascular tissues suggesting a function in amino acids export from the phloem into sink tissue
Solyc08g078190	no	0.83	**SGN**: Ethylene responsive transcription factor 1a (AHRD V1 *-*- C0J9I9_9ROSA); contains Interpro domain(s) IPR001471 Pathogenesis-related transcriptional factor and ERF, DNA-binding; **Pfam**: AP2 domain	AT4G17490 (BLAST in TAIR)	Encodes a member of the ERF (ethylene response factor) subfamily B-3 of ERF/AP2 transcription factor family (ATERF-6). The protein contains one AP2 domain. There are 18 members in this subfamily including ATERF-1, ATERF-2, AND ATERF-5. It is involved in the response to reactive oxygen species and light stress
				AT1G28360	**TAIR**: encodes a member of the ERF (ethylene response factor) subfamily B-1 of ERF/AP2 transcription factor family (ERF12). The protein contains one AP2 domain. There are 15 members in this subfamily including ATERF-3, ATERF-4, ATERF-7, and leafy petiole
				AT3G15210	**TAIR**: Encodes a member of the ERF (ethylene response factor) subfamily B-1 of ERF/AP2 transcription factor family (ATERF-4). The protein contains one AP2 domain. Acts as a negative regulator of JA-responsive defense gene expression and resistance to the necrotrophic fungal pathogen Fusarium oxysporum and antagonizes JA inhibition of root elongation. The mRNA is cell-to-cell mobile.
				AT5G51190	**TAIR**: encodes a member of the ERF (ethylene response factor) subfamily B-3 of ERF/AP2 transcription factor family. The protein contains one AP2 domain. There are 18 members in this subfamily including ATERF-1, ATERF-2, AND ATERF-5
				AT1G50640	**TAIR**: encodes a member of the ERF (ethylene response factor) subfamily B-1 of ERF/AP2 transcription factor family (ATERF-3). The protein contains one AP2 domain. There are 15 members in this subfamily including ATERF-3, ATERF-4, ATERF-7, and leafy petiole
				AT1G49120	**TAIR**: encodes a member of the ERF (ethylene response factor) subfamily B-6 of ERF/AP2 transcription factor family. The protein contains one AP2 domain. There are 12 members in this subfamily including RAP2.11.
				AT1G04370	**TAIR**: encodes a member of the ERF (ethylene response factor) subfamily B-3 of ERF/AP2 transcription factor family. The protein

					contains one AP2 domain. There are 18 members in this subfamily including ATERF-1, ATERF-2, AND ATERF-5
				AT3G20310	**TAIR**: Encodes a member of the ERF (ethylene response factor) subfamily B-1 of ERF/AP2 transcription factor family (ATERF-7). The protein contains one AP2 domain. Phosphorylated by PKS3 in vitro. Involved in ABA-mediated responses. Acts as a repressor of GCC box–mediated transcription together with AtSin3 and HDA19
				AT5G13910	**TAIR**: Encodes a member of the ERF (ethylene response factor) subfamily B-1 of ERF/AP2 transcription factor family (LEAFY PETIOLE). The protein contains one AP2 domain. There are 15 members in this subfamily including ATERF-3, ATERF-4, ATERF-7, and LEAFY PETIOLE. Acts as a positive regulator of gibberellic acid-induced germination
				AT4G34410	**TAIR**: encodes a member of the ERF (ethylene response factor) subfamily B-3 of ERF/AP2 transcription factor family. The protein contains one AP2 domain. There are 18 members in this subfamily including ATERF-1, ATERF-2, AND ATERF-5
				AT5G61600	**TAIR**: encodes a member of the ERF (ethylene response factor) subfamily B-3 of ERF/AP2 transcription factor family. The protein contains one AP2 domain. There are 18 members in this subfamily including ATERF-1, ATERF-2, AND ATERF-5
				AT5G44210	**TAIR**: encodes a member of the ERF (ethylene response factor) subfamily B-1 of ERF/AP2 transcription factor family (ATERF-9). The protein contains one AP2 domain. There are 15 members in this subfamily including ATERF-3, ATERF-4, ATERF-7, and leafy petiole.
				AT1G03800	**TAIR**: encodes a member of the ERF (ethylene response factor) subfamily B-1 of ERF/AP2 transcription factor family (ATERF-10). The protein contains one AP2 domain. There are 15 members in this subfamily including ATERF-3, ATERF-4, ATERF-7, and leafy petiole
				AT1G53170	**TAIR**: encodes a member of the ERF (ethylene response factor) subfamily B-1 of ERF/AP2 transcription factor family (ATERF-8). The protein contains one AP2 domain. There are 15 members in this subfamily including ATERF-3, ATERF-4, ATERF-7, and leafy petiole
				AT1G28370	**TAIR**: encodes a member of the ERF (ethylene response factor) subfamily B-1 of ERF/AP2 transcription factor family. The protein contains one AP2 domain. There are 15 members in this subfamily including ATERF-3, ATERF-4, ATERF-7, and leafy petiole
Solyc01g110350	no	0.82	**SGN**: Unknown Protein (AHRD V1)	AT4G38980	**TAIR**: Putative uncharacterized protein AT4g38980
Solyc04g040190	yes	-0.80	**SGN**: Lycopene beta-cyclase; **Pfam**: Lycopene cyclase protein	AT3G10230 (BLAST in TAIR)	**TAIR**: Encodes a protein with lycopene β-cyclase activity. This enzyme uses the linear, symmetrical lycopene as substrate. However, unlike the ε-cyclase which adds only one ring, the β-cyclase introduces a ring at both ends of lycopene to form the bicyclic β-carotene
				AT3G24200	**TAIR**: Ubiquinone biosynthesis monooxygenase COQ6, mitochondrial

Solyc04g057800	no	0.83	**SGN**: HAT family dimerisation domain containing protein (AHRD V1 ***- Q2R149_ORYSJ)	-	-
Solyc04g056480	no	0.89	**SGN**: Unknown Protein (AHRD V1)	-	-
Solyc04g016420	yes	0.88	**SGN**: Unknown Protein (AHRD V1)	AT5G04000	TAIR: Uncharacterized protein

Other candidate genes were also found in the network constructed from genes sharing an eQTL in chromosome 5. In this network, two genes, Solyc05g042180 and Solyc055014090, were presumed as candidate genes due to their high connection to other genes with trans-eQTL. In Sol Genomics Network, Solyc05g042180 is described as a nucleic acid binding protein (Table 5). This gene is orthologue to AT4G26000 in Arabidopsis, which is a protein that has an RNA-binding domain that regulates the response to short- and long-day photoperiodism. Solyc05g042180 is co-expressed with 13 other genes in the network. One of these genes, Solyc05g025510, is an orthologue to AT5G08560 that has a biological process of response to light stimulus (TAIR). Another co-expressed gene, Solyc05g013970, also has an RNA-binding domain just like Solyc05g042180. Furthermore, three genes (Solyc05g025510, Solyc05g014090 & Solyc05g025630) has a protein-binding domain and two of them are bind specific to histone (Solyc05g014090 & Solyc05g025630). Moreover, Solyc05g014090 is another possible candidate gene from the network explaining QTL for Gmax in chromosome 5. This gene is one of the high degree genes in the network, even though it is lower than Solyc05g042180. These two genes nevertheless are co-expressed together. Solyc05g014090 itself is described as a histone-binding protein. This gene is connected too with the other candidate gene in this network, Solyc05g042180.

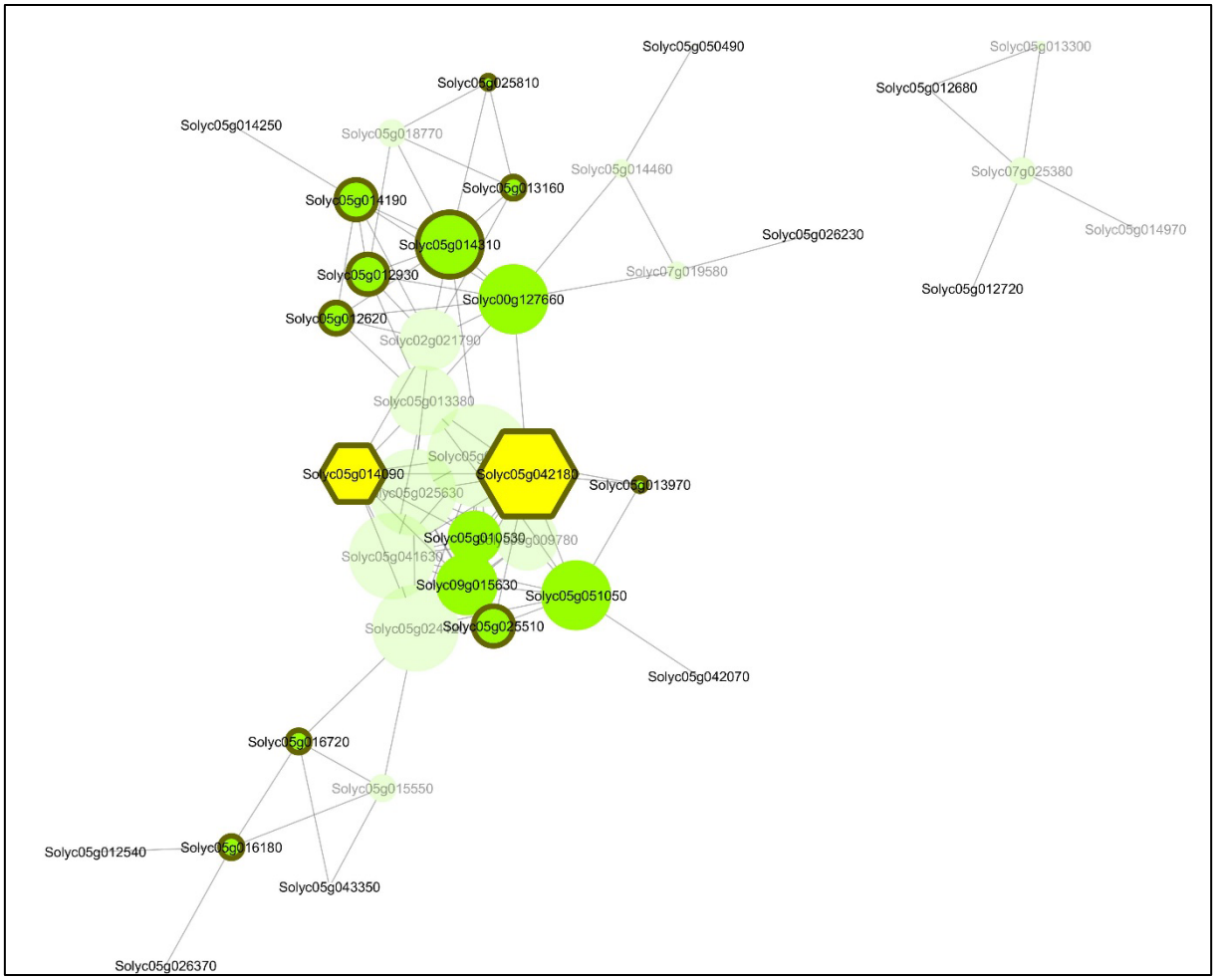


Figure 5. Network analysis for 61 genes located in the QTL in chromosome 5 for Gmax in LN maternal environment (peak locus: 6,814,373 mb). Genes are represented by nodes. Nodes with borders are genes located in the support interval of the QTL in chromosome 5. Larger nodes have more degree of connection compared to smaller nodes. Moreover, nodes with green color represent genes that have higher expression (or lower if it is an inhibitor) in dried PIM seed under LN maternal environment compared to the other genotype and maternal environment. These genes are connected to each other by edges. The candidate genes for this QTL is Solyc05g042180 and Solyc055014090, which is represented by a yellow hexagon-shaped node that is connected to more nodes representing gene with trans-eQTL. Note that several insignificant nodes and edges are not shown in the figure to maintain the image resolution.

Table 5. Description of Solyc05g042180 as a candidate gene for Gmax under LN maternal environment in chromosome 5 and its orthologs in Arabidopsis.

candidate gene

Gene	cis-eQTL	degree	description	orthologue in Arabidopsis	description
Solyc05g042180	yes	13	Nucleic acid binding protein (AHRD V1 *--- B6TJ76_MAIZE); contains Interpro domain(s) IPR018111 K Homology, type 1, subgroup	AT4G26000	Encodes a novel Arabidopsis gene encoding a polypeptide with K-homology (KH) RNA-binding modules, which acts on vegetative growth and pistil development. Genetic studies suggest that PEP interacts with element(s) of the CLAVATA signalling pathway
co-expressed genes					

Gene	cis-eQTL	correlation coefficient	description	ortholog in Arabidopsis	description
Solyc05g013970	yes	0.92	**SGN**: RNA-binding protein 39 (AHRD V1 *--- B5X2I3_SALSA); contains Interpro domain(s) IPR012677 Nucleotide-binding, alpha-beta plait; **Pfam**: RNA recognition motif	-	-
Solyc09g015630	no	0.91	Chlorophyllide a oxygenase, chloroplastic (AHRD V1 ***-CAO_ORYSJ)	-	-
Solyc05g041630	no	0.96	Genomic DNA chromosome 3 P1 clone MTO24 (AHRD V1 *---Q9LJA4_ARATH)	-	-
Solyc05g014090	yes	0.95	Histone-binding protein RBBP7 (AHRD V1 **-* B5X3G4_SALSA); contains Interpro domain(s) IPR017986 WD40 repeat, region; Pfam: WD domain, G-beta repeat	AT2G19540	Transducin family protein / WD-40 repeat family protein
Solyc05g025630	no	0.94	Histone-binding protein RBBP7 (AHRD V1 **-* B5X3G4_SALSA); contains Interpro domain(s) IPR017986 WD40 repeat, region; Pfam: Histone-binding protein RBBP4 or subunit C of CAF1 complex	AT2G19540	Putative WD-40 repeat protein
Solyc05g010530	no	0.93	Serine/threonine protein kinase family protein (AHRD V1 **** C6ZRM5_SOYBN); contains Interpro domain(s) IPR002290 Serine/threonine protein kinase; Pfam: Protein kinase domain	AT1G69910	Protein kinase superfamily protein
Solyc05g024120	no	0.93	Vacuolar protein sorting-associated protein 33 homolog (AHRD V1 ***- VPS33_ARATH)	-	-
Solyc05g025510	yes	0.91	**SGN**: WD-repeat protein-like (AHRD V1 ***-Q9FNN2_ARATH); contains Interpro domain(s) IPR017986 WD40 repeat, region; **Pfam**: WD domain, G-beta repeat	AT5G08560	WRDR26 is a WD-40 repeat containing protein initially identified as an interacting partner RanBPM. Its expression is induced by abiotic stress as well as various plant growth regulators including IAA, ABA and ethylene
Solyc00g127660	no	0.91	Unknown Protein (AHRD V1)	-	-
Solyc05g051050	no	0.90	**SGN**: Glycogen synthase kinase (AHRD V1 **** Q1AMT7_ORYSJ); contains Interpro domain(s) IPR002290 Serine/threonine protein kinase; **Pfam**: Protein kinase domain	AT3G61160	Protein kinase superfamily protein
				AT4G00720	Encodes ASKtheta, a group III Arabidopsis GSK3/shaggy-like kinase. Functions in the brassinosteroid signalling pathway
Solyc05g009780	no	-0.91	**SGN**: Methionine aminopeptidase (AHRD V1 ***-B9SEJ2_RICCO); contains Interpro domain(s) IPR002467 Peptidase M24A, methionine aminopeptidase, subfamily 1; Metallopeptidase family M24	AT1G13270	Encodes a methionine aminopeptidase formerly called MAP1B, renamed to MAP1C
				AT3G25740	Encodes a plastid localized methionine aminopeptidase. Formerly called MAP1C, now called MAP1B
				AT3G25740	Encodes a plastid localized methionine aminopeptidase. Formerly called MAP1C, now called MAP1B
Solyc05g012600	no	-0.95	Unknown Protein (AHRD V1)	AT2G03420	hypothetical protein
Solyc05g013380	no	-0.96	**SGN**: Alanine aminotransferase 2 (AHRD V1 **** A8IKE5_SOYBN); contains Interpro domain(s) IPR004839 Aminotransferase, class I and II; **Pfam**: Aminotransferase class I and II	AT1G70580	Encodes a protein with glyoxylate aminotransferase activity. It can act on a number of different small substrates and amino acids in vitro
				AT1G23310	Identified by cloning the gene that corresponded to a purified protein having glyoxylate aminotransferase activity. Localized to the peroxisome and thought to be involved in photorespiration/ metabolic salvage pathway

Table 6. Description of Solyc05g014090 as a candidate gene for Gmax under LN maternal environment in chromosome 5 and its orthologs in Arabidopsis.

candidate gene

Gene	cis-eQTL	degree	description	orthologues in Arabidopsis	description
Solyc05g014090	yes	9	Histone-binding protein RBBP7 (AHRD V1 **-* B5X3G4_SALSA); contains Interpro domain(s) IPR017986 WD40 repeat, region	AT2G19540	Transducin family protein / WD-40 repeat family protein

co-expressed genes

Gene	cis-eQTL	correlation coefficient	description	orthologues in Arabidopsis	description
Solyc05g025630	no	0.97	Histone-binding protein RBBP7 (AHRD V1 **-* B5X3G4_SALSA); contains Interpro domain(s) IPR017986 WD40 repeat, region	AT2G19540	Transducin family protein / WD-40 repeat family protein
Solyc05g041630	no	0.97	Genomic DNA chromosome 3 P1 clone MTO24 (AHRD V1 *---- Q9LJA4_ARATH)	-	-
Solyc05g010530	no	0.95	Serine/threonine protein kinase family protein (AHRD V1 **** C6ZRM5_SOYBN); contains Interpro domain(s) IPR002290 Serine/threonine protein kinase	AT1G69910	Protein kinase superfamily protein
Solyc05g012600	no	-0.91	Unknown Protein (AHRD V1)	AT2G03420	hypothetical protein
Solyc05g013380	no	-0.91	Alanine aminotransferase 2 (AHRD V1 **** A8IKE5_SOYBN); contains Interpro domain(s) IPR004839 Aminotransferase, class I and II	AT1G23310	Identified by cloning the gene that corresponded to a purified protein having glyoxylate aminotransferase activity. Localized to the peroxisome and thought to be involved in photorespiration/ metabolic salvage pathway
				AT1G70580	Encodes a protein with glyoxylate aminotransferase activity. It can act on a number of different small substrates and amino acids in vitro
Solyc05g024120	no	0.93	Vacuolar protein sorting-associated protein 33 homolog (AHRD V1 ***- VPS33_ARATH)	-	-
Solyc09g015630	no	0.91	Chlorophyllide a oxygenase, chloroplastic (AHRD V1 ***-CAO_ORYSJ)	-	-
Solyc02g021790	no	0.90	Unknown Protein (AHRD V1)	AT2G01050	zinc ion binding / nucleic acid binding protein
Solyc05g042180	yes	0.95	Nucleic acid binding protein (AHRD V1 *--- B6TJ76_MAIZE); contains Interpro domain(s) IPR018111 K Homology, type 1, subgroup	AT4G26000	Encodes a novel Arabidopsis gene encoding a polypeptide with K-homology (KH) RNA-binding modules, which acts on vegetative growth and pistil development. Genetic studies suggest that PEP interacts with element(s) of the CLAVATA signaling pathway

Network in chromosome 6 also yields in a possible candidate gene for AUC in LN (Figure 6). In fact, the cluster analysis results in the construction of two gene cluster for this trait. However, the high degree node from one of the cluster represented a gene that is neither differentially expressed in both parental and maternal environment, nor is collocated in the support interval of the phenotypic QTL. Therefore, I

exclude this gene from being considered as a candidate gene for AUC in LN. Meanwhile, on the other cluster, there is a high degree node that is also located within the interval range of the phenotypic QTL, which fulfills criteria as a candidate gene. This gene was identified as Solyc0609710 and connected to the other seven genes in the network. Interestingly, this gene is also present in the network for Gmax based on the QTL in the chromosome 4 (Figure 4).

Solyc0609710 is connected to the other seven genes that are located on chromosome 6 even though not in the support interval of the phenotypic QTL. However, these genes have various functions and none seems related to seed germination or light response, including their orthologues in Arabidopsis. Solyc0609710 itself has many orthologues genes in Arabidopsis. One of the most similar genes based on the sequence alignment is AT5G39610. This gene is a NAC-domain transcription factor that is involved in the regulation of seed germination, which provides another evidence for Solyc0609710 as the candidate genes for AUC.

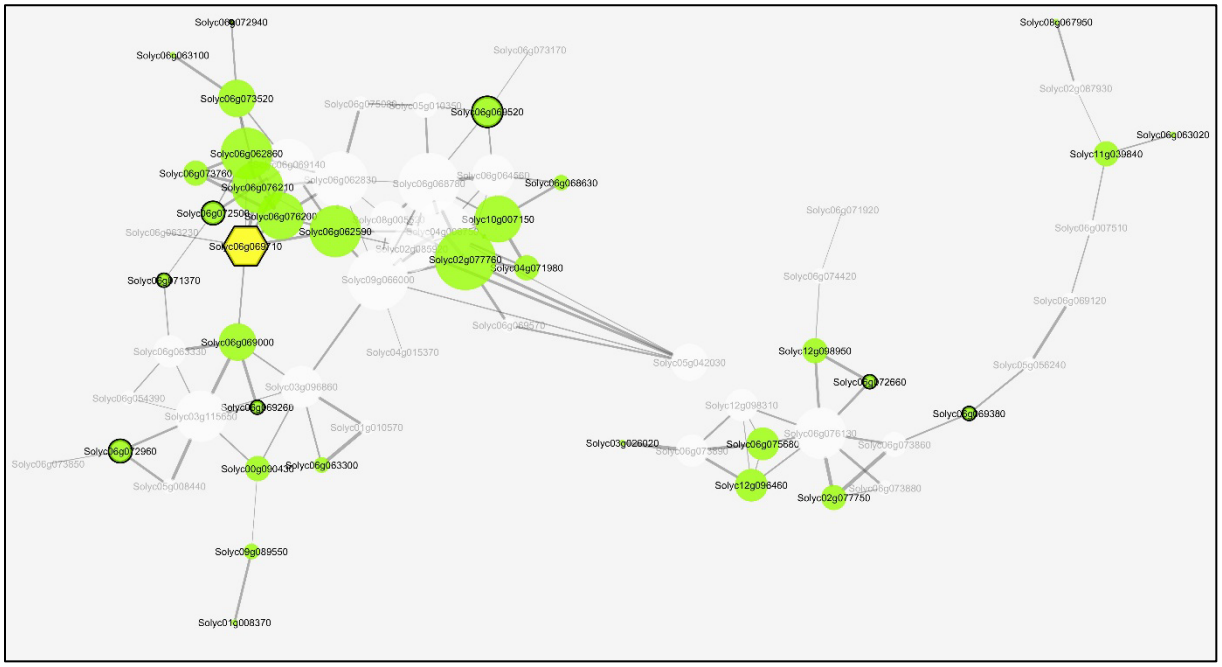


Figure 6. Network analysis for 75 genes located in the QTL in chromosome 6 for AUC in LN maternal environment (peak locus: 43,761,285 mb). Genes are represented by nodes. Nodes with borders are genes located in the support interval of the QTL on chromosome 6. Larger nodes have more degree of connection compared to smaller nodes. Moreover, nodes with green color represent genes that have a higher expression (or lower if it is an inhibitor) in dried MM seed under LN maternal environment compared to the other genotype and maternal environment. These genes are connected to each other by edges. Thick edges indicate strong correlation and the other way around for thin edges. The candidate genes for this QTL is Solyc0609710, which is represented by a yellow hexagon-shaped node that is connected to the other seven nodes that have trans-eQTL in the phenotypic QTL region. Note that several insignificant nodes and edges are not shown in the figure to maintain the image resolution.

Table 7. Description of Solyc06g069710 as a candidate gene for Gmax under LN maternal environment in chromosome 6 and its orthologs in Arabidopsis.

candidate gene

Gene	cis-eQTL	connection degree	description	ortholog in Arabidopsis	description
Solyc06g069710	yes	7	NAC domain protein IPR003441 (AHRD V1 ***-B9I557_POPTR); contains Interpro domain(s) IPR003441 No apical meristem (NAM) protein	AT5G39610	Encodes a NAC-domain transcription factor. Positively regulates aging-induced cell death and senescence in leaves. This gene is upregulated in response to salt stress in wildtype as well as NTHK1 transgenic lines although in the latter case the induction was drastically reduced. It was also upregulated by ABA, ACC and NAA treatment, although in the latter two cases, the induction occurred relatively late when compared with NaCl or ABA treatments
				AT5G53950	Transcriptional activator of the NAC gene family, with CUC1 redundantly required for embryonic apical meristem formation, cotyledon separation and expression of STM
				AT2G24430	NAC domain containing protein 38
				AT3G04060	NAC046 is a member of the NAC domain containing family of transcription factors
				AT3G29035	Encodes a protein with transcription factor activity
				AT1G76420	Identified in an enhancer trap line; member of the NAC family of proteins. Expressed at the boundary between the shoot meristem and lateral organs and the polar nuclei in the embryo sac.
				AT5G18270	NAC domain containing protein 87
				AT4G28530	NAC domain containing protein 74
				AT3G15170	Encodes a transcription factor involved in shoot apical meristem formation and auxin-mediated lateral root formation. The gene is thought not to be involved in stress responses (NaCl, auxins, ethylene).

co-expressed gene

Gene	cis-eQTL	connection degree	description	ortholog in Arabidopsis	description
Solyc06g063230	no	-0.86	Unknown Protein (AHRD V1)	AT1G71010 (BLAST in TAIR)	Encodes a protein that is predicted to act as a phosphatidylinositol-3P 5-kinase, but, because it lacks a FYVE domain, it is unlikely to be efficiently targeted to membranes containing the proposed phosphatidylinositol-3P substrate. Therefore, its molecular function remains unknown. The mRNA is cell-to-cell mobile.
				AT1G73090	WD repeat protein

Solyc06g076200	no	0.92	X1 (Fragment) (AHRD V1 ***- Q3T7E5_ZEAMP); contains Interpro domain(s) IPR005380 Region of unknown function XS	AT1G13790	Belongs to a subgroup of SGS3-like proteins that act redundantly in RNA-directed DNA methylation
Solyc06g076210	no	0.92	X1 (Fragment) (AHRD V1 ***- Q3T7E5_ZEAMP); contains Interpro domain(s) IPR005379 Region of unknown function XH	AT1G13790	Belongs to a subgroup of SGS3-like proteins that act redundantly in RNA-directed DNA methylation
Solyc06g062590	no	0.90	Transmembrane 9 superfamily protein member 4 (AHRD V1 **** B6SXZ2_MAIZE); contains Interpro domain(s) IPR004240 Nonaspanin (TM9SF)	AT1G55130	Encodes an Arabidopsis Transmembrane nine (TMN) protein. Transmembrane nine (TM9) proteins are localized in the secretory pathway of eukaryotic cells and are involved in cell adhesion and phagocytosis.
	no			AT3G13772	Encodes an Arabidopsis Transmembrane nine (TMN) protein. Transmembrane nine (TM9) proteins are localized in the secretory pathway of eukaryotic cells and are involved in cell adhesion and phagocytosis. Overexpression of this protein in yeast alters copper and zinc homeostasis.
Solyc06g062860	no	0.86	Ataxin-2 (AHRD V1 *--- D3BPE8_POLPA); contains Interpro domain(s) IPR009604 LsmAD domain	AT1G54170	ataxin-2-related, similar to SCA2 (GI:1770390) (Homo sapiens); similar to ataxin-2 (GI:3005020) (Mus musculus). Member of a family of PAM2 motif containing proteins.
	no			AT3G14010	hydroxyproline-rich glycoprotein family protein, similar to Mrs16p (GI:2737884) (Saccharomyces cerevisiae); weak similarity to ataxin-2 related protein (GI:1679686) (Homo sapiens). Included in a family of CTC interacting domain proteins found to interact with PAB2.
Solyc06g069000	no	-0.87	Uncharacterized GPI-anchored protein At4g28100 (AHRD V1 ***- UGPI7_ARATH)	AT4G39400	Encodes a plasma membrane localized leucine-rich repeat receptor kinase involved in brassinosteroid signal transduction. BRI1 ligand is brassinolide which binds at the extracellular domain. Binding results in phosphorylation of the kinase domain which activates the BRI1 protein leading to BR responses.
				AT3G18050	GPI-anchored protein
Solyc06g069140	no	0.89	Pre-rRNA processing protein (AHRD V1 *-*-A8JCZ5_CHLRE); contains Interpro domain(s) IPR005304 Ribosomal biogenesis, methyltransferase, EMG1/NEP1	AT1G28640	GDSL-motif esterase/acyltransferase/lipase. Enzyme group with broad substrate specificity that may catalyze acyltransfer or hydrolase reactions with lipid and non-lipid substrates.
				AT3G57000	nucleolar essential protein-like protein

3.6. Validation of the candidate gene expression in germinating seed

As previously mentioned, the eQTL data used for the identification of the candidate genes were obtained from dried seeds. Therefore, the expression of the candidate genes needs to be validated in the germinating seeds. This analysis is also necessary to see whether there are differences in gene expression among parental lines as a criterion of the candidate gene determination. To do this, the gene expression was quantified using two-step qPCR assay (Kubista et al. 2006). This kind of assay is very sensitive to the upstream process, such as RNA isolation, cDNA synthesis, and others (Fleige & Pfaffl 2006). Therefore, only good samples were used in this experiment.

In order to determine its quality, the integrity of RNA was assessed on a gel electrophoresis, and the OD260/OD280 and concentration were measured using Trinean Xpose spectrophotometer (BIOKÉ). All samples used for cDNA synthesis showed two bands on gel electrophoresis, which represented the intact 18s and 28s rRNAs (the figure is not shown). Several samples also displayed the presence of a smear and genomic DNA, but they were disappear after DNAse treatment. Moreover, the concentration of the RNA and OD260/OD280 in most samples were higher than 100 ng/ul and in between 1.9—2, respectively (Table 8). Only two sample had bad quality of RNA, which was sample number 3 and number 24, yet the cDNA was synthesized anyway using the available samples. The synthesized cDNA was served as a template for qPCR.

Table 8. The concentration and OD260/OD280 of RNA samples

Sample name	RNA (ng/ul)	A260/A280	Sample name	RNA (ng/ul)	A260/A280
1	208	2.02	31	165.1	2.03
2	181.4	2.08	32	209.3	2.2
3	62	1.92	33	113.2	2.01
4	120.9	2.02	34	118.1	1.96
5	144	2.08	35	208.4	2.04
6	125.9	1.97	36	143.4	2.04
7	123.6	1.9	37	152.9	2.03
8	144.8	2.09	38	120	2.02
9	206.8	2.06	39	118	1.97
10	150	1.91	40	178.3	2.07
11	173.5	2.06	41	137.6	1.99
12	140.6	1.95	42	182.4	1.99
13	163.9	1.87	43	198.4	2.09
14	169.9	2.04	44	143.6	2.01
15	122.5	2.05	45	121	2.08
16	172.1	2.08	46	125.3	2.03
17	145.1	1.9	47	159.3	2.03
18	134.2	2.07	48	353.8	2.03
19	31.6	1.82	49	124.5	2.04
20	98.6	2.08	50	147.9	2.08

21	164.3	2.06	51	125.6	1.88
22	132.1	1.95	52	181.2	2.06
23	177.9	2.08	53	112.5	2.08
24	-	1.82	54	170	2.03
25	141.3	2.09	55	157.4	2.02
26	129.4	2.02	56	176.7	2.08
27	204.1	1.99	57	149.4	2.1
28	167.6	2.04	58	188.2	1.98
29	153	2.09	59	263.3	2.04
30	157.9	2.04	60	142.3	1.88

The primers for each of the candidate gene was designed using CLC Main Workbench 7 before the qPCR assay. Next, the efficiency of these primers was assessed on diluted cDNA samples by plotting the linear regression of the ct value. The result can be seen on Appendix 6—11. The primers used in this experiment has an efficiency ranging from 100—110%, except for Solyc04g014090 where the efficiency is only 83% yet the experiment was continued anyway. Another obstacle was the primer for Solyc05g042180 did not produce any amplicon at all. This could be an indication that the gene is not expressed in seed. However, it does not seem to be the case since the gene is expressed in the dried seed based on the RNA-seq data. This also could happen because the gene sequence difference between the genotype under study and MM cv. Heinz, which is used as the reference gene for primer design. If this was the case, then it would be obvious if no amplicon was produced since the primer could not anneal to any cDNA sequence. Aside from this, the expression analysis for the candidate gene was continued using the other well-working primers.

The qPCR assay for each of the candidate gene was performed. The Livak method was used to analyze the data and the expression of the dried seed of MM in HP was served as the calibrator. The result was shown in such way so that the relative expression in both genotype and maternal environment could be easily compared (Figure 7-9).

The relative expression of Solyc05g014090 is shown in Figure 7. This gene was expected to be involved in the phenotypic variation for Gmax particularly in LN because the QTL for this trait is only present in LN but not in HP. The positive allele for this gene was assumed coming from PIM due to the negative effect of the QTL. As can be seen in Figure 7, the expression of the gene was very high in the dried seed of all treatment and decreasing on 6 hai to 12 hai. However, this was not occurred in PIM under LN maternal environment. In this treatment, the expression was already low in dried seed without hardly any changes on 6 hai and 12 hai. Moreover, in 12 hai, there was not any difference between the gene expression in

light and dark condition. This is in contrary with the result of the other samples where the dark-treated seed had higher expression than light-treated seed in 12 hai.

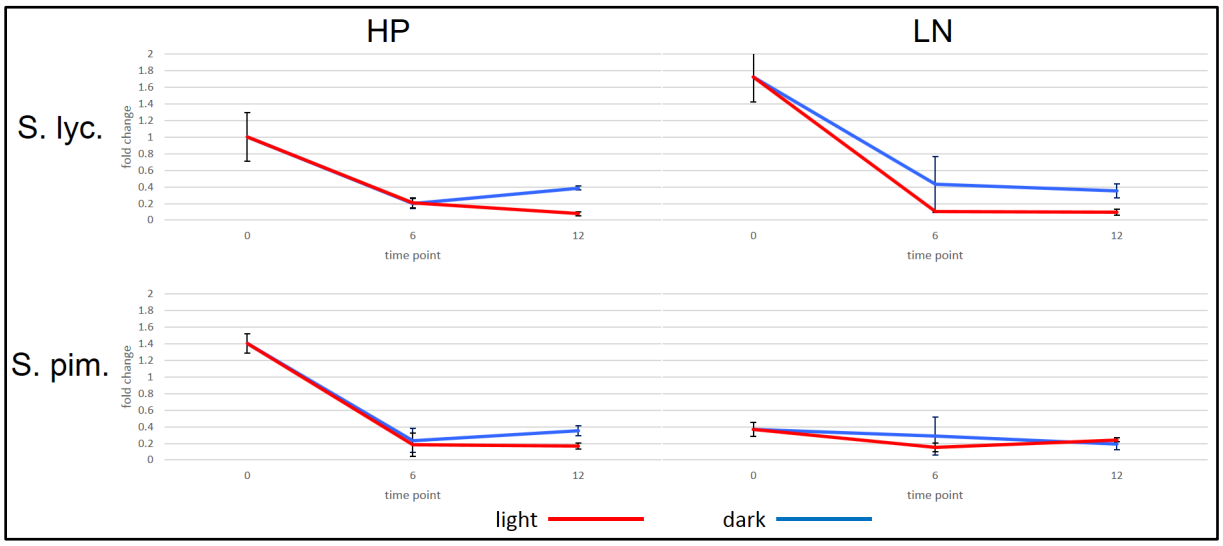


Figure 7. The mean expression ± standard error of Solyc05g014090 of three independent biological replicates in dried seed, 6 hours after imbibition, and 12 hours after imbibition across different genotype, maternal environment, and light condition. The expression value was determined relative to the expression in the dried seed of Solanum lycopersicum in HP. S. lyc: Solanum lycopersicum cv. Moneymaker; S. pim: Solanum pimpinellifolium; HP: HP maternal environment; LN: LN maternal environment. Red line: average expression in light condition. Blue line: average expression in dark condition.

Figure 8 shows the comparison of Solyc04g051720 gene expression across different genotype and maternal environment. This gene is the possible candidate gene for Gmax under HP maternal environment based on the network analysis. The QTL for this trait had a positive effect, indicating that the positive allele was originated from MM cv. Moneymaker. In Figure 8, the gene was scarcely expressed in dried seed as well as in germinating seed of MM until 12 hai. In contrary, the regulation of Solyc04g051720 was presented in PIM, particularly in dried seed where the expression was displayed as the highest. Furthermore, dried PIM seed in HP had two times higher expression of the gene compared to the dried seed in LN. The expression in both maternal environments appeared to be decreasing drastically during seed imbibition yet it was still relatively higher compared to the expression in MM.

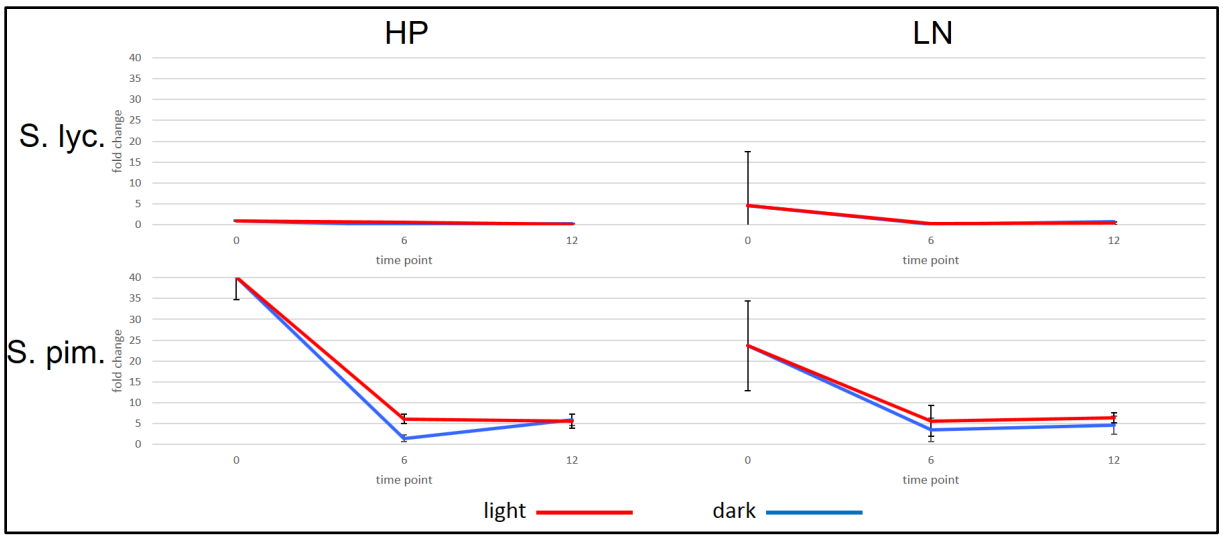


Figure 8. The mean expression ± standard error of Solyc04g051720 of three independent biological replicates in dried seed, 6 hours after imbibition, and 12 hours after imbibition across different genotype, maternal environment, and light condition. The expression value was determined relative to the expression in the dried seed of Solanum lycopersicum in HP. S. lyc: Solanum lycopersicum cv. Moneymaker; S. pim: Solanum pimpinellifolium; HP: HP maternal environment; LN: LN maternal environment. Red line: average expression in light condition. Blue line: average expression in dark condition.

The expression of Solyc06g069710 in dried and imbibed seed under different genotype and the maternal environment was shown in Figure 9. This gene was presumed as a candidate gene for AUC in LN environment. Due to positive QTL effect, it was assumed that the source of the positive allele for the trait under study is from MM. Based on the qPCR result, the expression of Solyc06g069710 in the dried seed of MM was two-fold higher than in dried seed of PIM. In the latter species, the expression was decreasing in 6 hai and 12 hai. However, in the other species, the expression of Solyc06g069710 were regulated distinctly under different light condition. This was clearly illustrated in Figure 9 where the seed imbibed in the dark had higher gene expression than in light condition. Moreover, the differences in the gene expression was evidenced to be occurred earlier in LN at 6 hai.

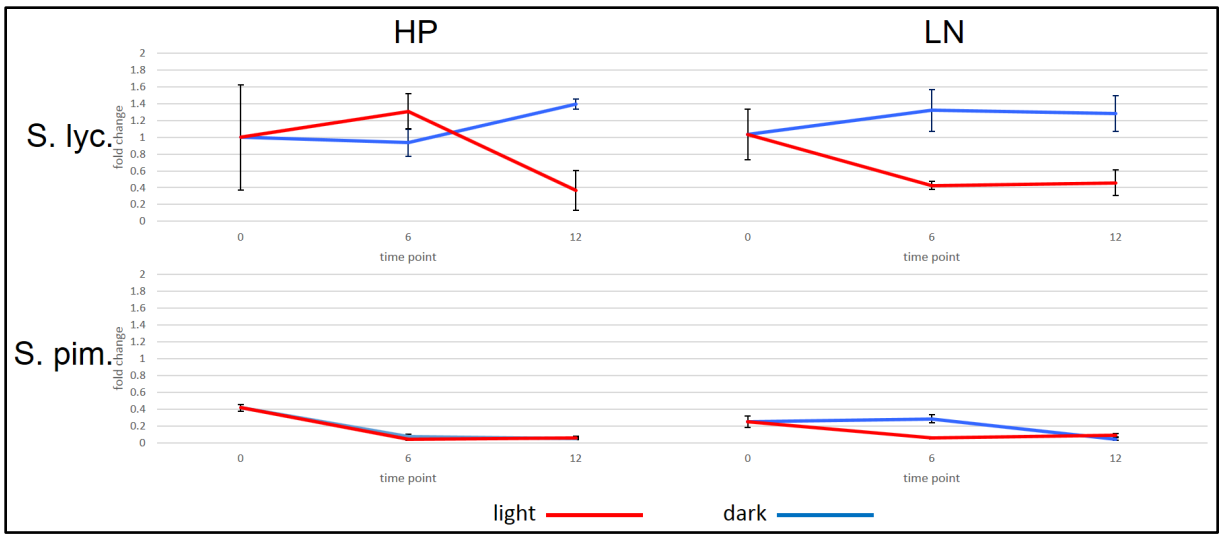


Figure 9. The mean expression ± standard error of Solyc06g069710 of three independent biological replicates in dried seed, 6 hours after imbibition, and 12 hours after imbibition across different genotype, maternal environment, and light condition. The expression value was determined relative to the expression in the dried seed of Solanum lycopersicum in HP. S. lyc: Solanum lycopersicum cv. Moneymaker; S. pim: Solanum pimpinellifolium; HP: HP maternal environment; LN: LN maternal environment. Red line: average expression in light condition. Blue line: average expression in dark condition.

3.7. Expression analysis of *PHYB2*, *PHYA*, and *GIGANTEA* in dried and germinating seed

Aside from three candidate genes, I also analyze the expression of *PHYB2*, *PHYA*, and *GIGANTEA* that are known to be involved in tomato seed germination. For *PHYB2*, the gene expression in dried and imbibed seed was presented in Figure 10. The gene expression in MM dark-treated seed was shown to be stable from dried seed until 12 hai with a slight reduction LN maternal environment. However, the result pointed out that the expression of *PHYB2* in light-treated seed was downregulated during imbibition and it took place earlier in LN environment. The same pattern was also happened in PIM. Interestingly, the *PHYB2* expression in dark-imbibed was also downregulated, leading to the same fold changes with light-treated seed. The downregulation of *PHYB2* expression in dark-imbibed seed, nevertheless, occurred later at 6 hai in LN than in HP.

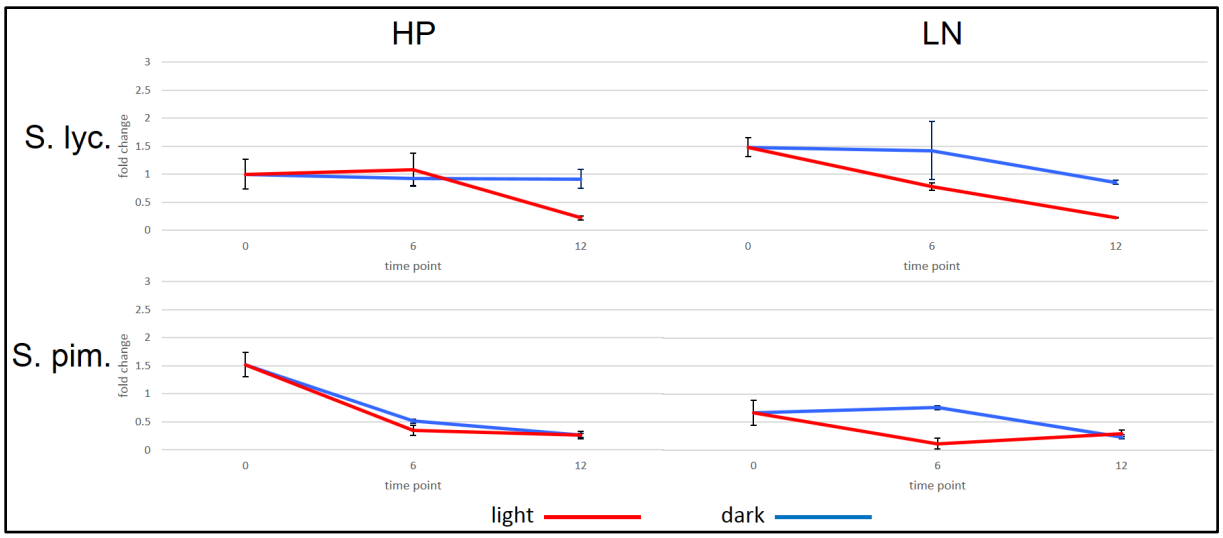


Figure 10. The mean expression ± standard error of *PHYB2* of three independent biological replicates in dried seed, 6 hours after imbibition, and 12 hours after imbibition across different genotype, maternal environment, and light condition. The expression value was determined relative to the expression in the dried seed of Solanum lycopersicum in HP. S. lyc: Solanum lycopersicum cv. Moneymaker; S. pim: Solanum pimpinellifolium; HP: HP maternal environment; LN: LN maternal environment. Red line: average expression in light condition. Blue line: average expression in dark condition.

The gene lower expression in light treatment is also observed in *PHYA*. As we can see in Figure 11, the gene expression as the response of dark imbibition, in general, was slightly increased from dried seed until 12 hai. A slight downregulation in 6 hai was only shown in PIM under HP yet it was increased again in 12 hai. This, however, did not occur in the light-treated seed. Except in MM under HP maternal environment, the gene expression of *PHYA* was decreasing at 6 hai and it was continued until 12 hai. Eventually, at 12 hai, the gene relative expression in all treatment was lower in light-treated seed in comparison to dark-treated seed.

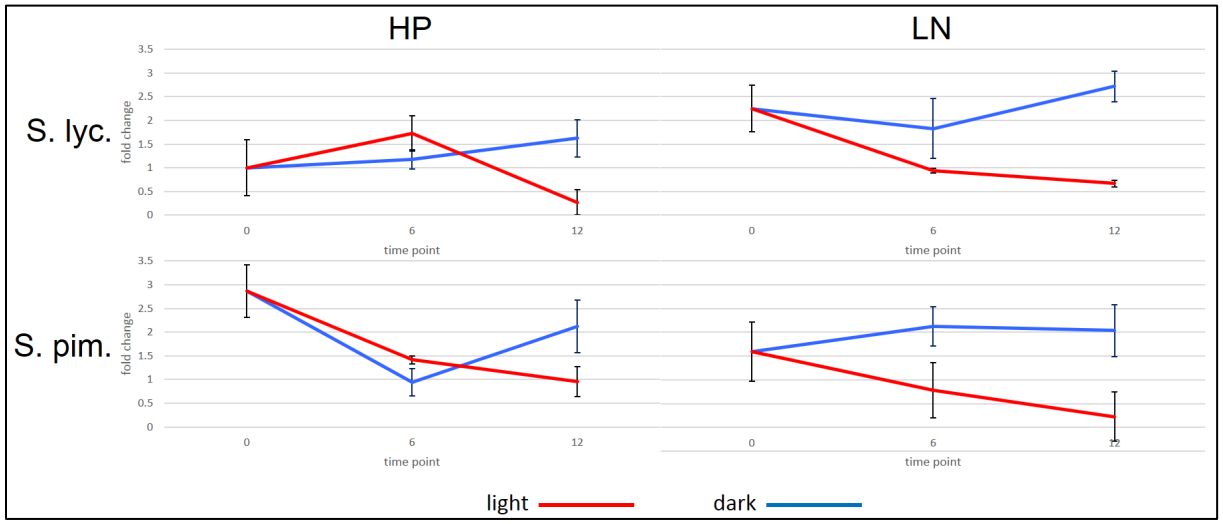


Figure 11. The mean expression ± standard error of *PHYA* of three independent biological replicates in dried seed, 6 hours after imbibition, and 12 hours after imbibition across different genotype, maternal environment, and light condition. The

expression value was determined relative to the expression in the dried seed of Solanum lycopersicum in HP. S. lyc: Solanum lycopersicum cv. Moneymaker; S. pim: Solanum pimpinellifolium; HP: HP maternal environment; LN: LN maternal environment. Red line: average expression in light condition. Blue line: average expression in dark condition.

The last gene expression analysis was for *GIGANTEA* that is presumed to be involved in light response in tomato seed germination. Based on the expression analysis in Figure 12, the gene expression in dried seed was higher in PIM under HP maternal environment compared to the other treatments. The gene expression, however, was decreasing when the seed was imbibed. At 6 hai, the expression of GI apparently was higher in light-imbibed seed compared to dark-imbibed seed in all genotype and environment except in PIM under LN maternal environment. Nevertheless, at 12 hai not much difference was observed between dark- and light imbibed seed in both species under different maternal environment.

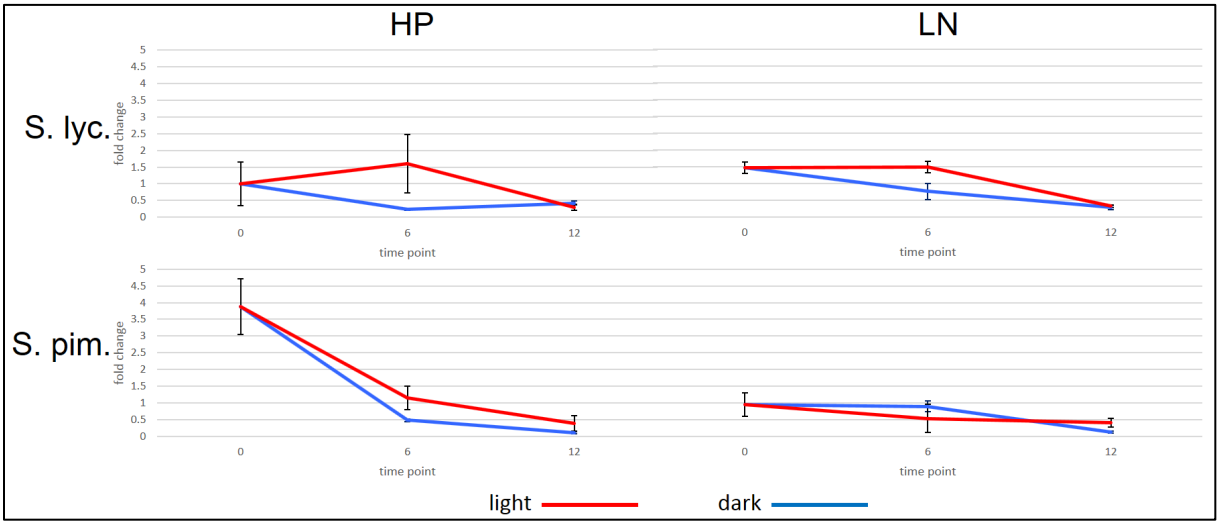


Figure 12. The mean expression ± standard error of *GIGANTEA* of three independent biological replicates in dried seed, 6 hours after imbibition, and 12 hours after imbibition across different genotype, maternal environment, and light condition. The expression value was determined relative to the expression in the dried seed of Solanum lycopersicum in HP. S. lyc: Solanum lycopersicum cv. Moneymaker; S. pim: Solanum pimpinellifolium; HP: HP maternal environment; LN: LN maternal environment. Red line: average expression in light condition. Blue line: average expression in dark condition.

4. DISCUSSION

4.1. The difference in seed germination onset and rate as the response to light in *Solanum lycopersicum* cv. Moneymaker and *Solanum pimpinellifolium*

As described previously in the result, the capability of seed to germinate—indicated by Gmax—was not affected by light. Instead, light exposure dictates the onset and the rate of germination. This was respectively indicated by the parameter of t50 and t10 in the experiment result. Interestingly, the two species under study behave differently, suggesting the different mechanism and role of light in the life cycle of both species. Even though this is not a priority trait in breeding, seed germination onset and rate eventually could be served as an indicator of the high-quality seedling. This is because longer germination rate is associated with a high percentage of the abnormal seedling establishment, at least in tomato (Tarquis & Bradford 1992) and lettuce (Bewley et al. 2013). Therefore, this part of the discussion will focus mainly on the seed germination rate as a response to light, particularly in the ecological relevance and physiological mechanism.

4.1.1. Ecological relevance

The seed of PIM had already been long known to have faster germination rate compared to its modern cultivated species (Whittington et al. 1965, Thompson 1974). This trait—which has been argued to be regulated by a common genetic mechanism (Foolad et al. 2007)—was observed under various stress and non-stress environment. The small dimension of PIM seed may be the cause of its faster germination rate. This argument is supported by studies in diesel tree (Souza & Fagundes 2014), barley (Gharoobi 2006), and several tropical tree species (Murali 1997), altogether demonstrating that smaller seeds germinate more rapidly compared to larger seeds. In tomato, however, genetic and environmental studies showed that the effect of seed mass to germination performance is rather small (Pet & Garretsen 1983) and there was no direct correlation between seed dimension and seed performance (Khan et al. 2012).

Even though it is not directly correlated with seed germination rate, seed mass is in fact associated with the light requirement for the initiation of seed germination. A germination study in neotropical pioneer species indicated that the requirement of light is increasing along with the decrease of seed size (Pearson et al. 2002). The correlation between seed size and light requirement could be explained because light may be the cue for small-seeded plants for an open gap in the tree canopy, hence the germination is initiated as soon as possible. This phenomenon was also observed on another experiment in woody

pioneer species of Amazon secondary forest (Aud & Ferraz 2012), temperate tree species in Japan (Xia et al. 2016), and succulent species in Mexico (Flores et al. 2016). These examples indicate that the light dependence for the small seed to initiate germination is a common mechanism across species and ecosystem. Therefore, the relation between seed size and light dependency could explain high germination rate of small-sized seed PIM under light treatment.

How about in MM? Although larger than PIM, the seed of MM is still considered small, thus light requirement should persist in such species (Pearson et al. 2002). However, my experimental result in both maternal environments showed that light was a suppressive factor in MM seed germination, indicated by lower t10, t50, and AUC (see Figure 1 and Figure 2). This, nevertheless, is an exceptional case in *Solanum* genus since the promoting effect of light for seed germination is common among the wild species of such taxa group, for instance in *Solanum ptycanthum* (Zhou et al. 2005), *Solanum aviculare* (Porter & Clark 1979), *Solanum laciniatum* (Porter & Clark 1979), and *Solanum hayesii* (Pearson et al. 2002). Hence, considering its status as a modern cultivated species, the loss of light-dependency for seed germination in MM may be caused by an intensive selection during the breeding effort. Still and all, not all breeding processes of tomato results in the loss of light-dependency because there is a domesticated tomato cultivar, namely *S. lycopersicum* cv. Taiwan Red, that still preserves light-dependency in seed germination (Gui et al. 1991).

Not only the loss in light requirement, the seed weight of MM is also increased as the result of domestication (Doganlar et al. 2000). For many neotropical pioneer seeds, the increase of seed weight and the decrease of the light requirement coincide with the thicker seed coat. As the result, larger seeds would have a higher probability of seedling establishment due to better protection against disease and predation (Pearson et al. 2002). Yet, whether the increase in seed weight of domesticated tomato is associated with the loss of light dependency is still in question.

One thing for sure is that light had become a suppressive environmental factor that inhibits seed germination of MM. This may be explained because of high irradiation response (HIR) due to excessive photon dose received by the seed, which is known to inhibit seed germination in many species (Baskin & Baskin 2014). Prolonged exposure to light can be an environmental cue for high solar radiation and dry condition that are harmful to plant growth (Bewley et al. 2013). That is why numerous species of plant developed HIR to prevent seed germination, including MM.

4.1.2. Physiological and molecular mechanisms

Little is known about the mechanism underlying seed germination rate as the response to light. The factors that regulate the difference in seed germination rate can be activated during any of the three phases of seed germination. For instance, in the imbibition phase, water potential can affect the germination rate (Bewley et al. 2013). However, this property is basically influenced more by the physical process and does not seems to be associated with light.

The role of light in seed germination rate may occurs during the second and third phase of seed germination. Particularly, Groot & Karssen (1987) argued that the germination rate is regulated by genes related to storage mobilization and endosperm weakening, which is under the influence of many environmental factors such as red light. This argument was supported by Eckstein et al. (2016) who discovered that galactomannan-hydrolyzing enzyme was detected earlier in the endosperm tip of fast germinating cultivar compared to slow germinating cultivar and they showed that the activity of this enzyme is inhibited by far-red light and induced by red light. Another study also showed that the increase of storage starch in germinating seed was higher under red light compared to far-red light treatment. In addition, the activity of isocitrate lyase and malate synthase that involve in gluconeogenesis during seed germination was also under the influence of light quality (Khan et al. 2012).

4.2. QTL analysis in the RIL population revealed positive alleles for seed germination traits originated from both parental lines

The germination assay was also performed on RIL population derived from the crossing of MM and PIM. The result showed that the mean of RIL population is in between the parental line for all trait under study (Figure 1 and Figure 2). Moreover, the mean for each trait was accompanied with a large standard error implying a wide genetic variation among RILs, which later confirmed by QTL analysis (see the explained variance in Table 1). The phenotype of the RIL was also distributed in an approximately normal distribution (see Figure 3) with the presence of transgressive progenies in both sides of the histogram in almost all traits studied, particularly in traits where both parents had more or less similar values.

The presence of transgressive progenies may be explained by the overdominance effect of the allele. This occurs when the heterozygous allele has a higher value compared to both homozygous alleles. However, the occurrence of overdominance effect is very unlikely in case of selfed F8 RIL population because low proportion or even the absence of heterozygous allele.

The transgressive phenotypes can also be caused by the accumulation of complementary alleles with the same direction from both parents into the inbred lines. For example, the presence of RILs with faster seed germination onset in LN—indicated by smaller t10 in Figure 1—could be caused by the inheritance of positive alleles for t10 from both parents. This, however, can only occur if both parents has a mixture of positive and negative alleles for a certain seed germination parameter regardless of its phenotypes. For instance, MM may have positive allele(s) that determine the positive direction for t10, even though the phenotype of such species had the opposite. If that is true, then we can also argue that there are different genes regulating traits where both parents had similar phenotypes, such as Gmax. This is because the occurrence of many RILs that have germination percentage much lower than the parents had (see Figure 3) which could be explained by the inheritance of many complementary genes that have a negative effect from both parents onto such traits. In fact, the emergence of RILs is more likely for this kind of trait, where both parents has similar phenotypes. This argument was proven by the negative association between the difference in the parental means and the appearance of transgressive progenies (DeVicente & Tanksley 1993).

The presence of allele with opposite effect from the parent phenotype seems counterintuitive. In Figure 1 and Figure 2, we can see that the seed germination performance of PIM appeared to be more superior, particularly in terms of t10, t50, u8416, and AUC, where the means of these values was much better compared to its domesticated species. Thus, one may assume that PIM was the source of all positive alleles for seed germination performance of the RILs population. This, however, is not completely true since several of the beneficial alleles were also discovered to be coming from MM from the result of QTL analysis. In Table 9, we can see that the positive alleles for seed germination quality were originated from both parents. Nevertheless, PIM had the most contribution of these alleles; only seven out of twenty-nine from MM. This suggests that the gene for light germination is still conserved in the domesticated species even after hundreds of years of breeding effort.

Table 9. quantitative trait locus for Gmax, t10, t50, u8416, and AUC in light germination.

Maternal environment	Trait	Light condition	Chr	Marker peak	Support interval (cM)	LOD	Explained variance	µ S. lyc.	µ het	µ S. pimp	Effects	Source
LN	AUC	light	1	76672459	99.7-119.8	2.26	5.1	36.1	41.2	46.3	-4.19	PIM
LN	t50	light	4	4936160	41.0-51.5	3	8.3	50.2	56.2	62.1	-5.96	MM
LN	t10	light	4	5460003	42.0-55.7	2.87	8.3	38	42.5	47	-4.42	MM
HP	Gmax	light	4	51678315	55.1-59.4	5.51	24.9	1.26	1.46	1.67	-0.2	PIM
LN	t50	light	4	55204967	67.3-73.6	3.05	8.4	51.4	57.4	63.3	-5.93	MM
HP	u8416	light	5	266350	0-8	2.43	10.6	645.6	656	666.5	-10.47	MM
LN	t10	light	5	416826	0-10.0	2.09	5.9	38.7	42.5	46.3	-3.78	MM
LN	t50	light	5	6711222	51.4-65.9	2.24	5.8	61.7	56.7	51.8	4.94	PIM
LN	Gmax	light	5	6814373	51.4-63.2	2.54	9.3	1.67	1.81	1.96	-0.14	PIM
HP	Gmax	light	5	60619277	66.9-83.4	3.01	10.4	1.33	1.46	1.59	-0.13	PIM
HP	Gmax	light	6	31371391	9.9-24.5	2.72	11.2	1.6	1.46	1.32	0.13	MM

LN	AUC	light	6	43761285	92.8-102.1	3.03	8.3	46.6	41.4	36.2	5.23	MM
HP	AUC	light	7	28076245	23.4-35.4	2.08	8.9	35.9	40.8	45.6	-4.86	PIM
HP	Gmax	light	7	28076245	26.4-36.4	2.35	9.6	1.34	1.46	1.59	-0.12	PIM
HP	t10	light	7	28076245	20.4-32.4	2.14	10.1	43.2	38.4	33.6	4.8	PIM
HP	t50	light	7	28076245	22.4-35.4	2.66	12.6	63.5	56	48.5	7.52	PIM
LN	u8416	light	7	28076245	24.4-39.8	2.57	8.4	28.7	25.6	22	3.33	PIM
HP	AUC	light	11	47543161	14.3-27.8	2.17	9.3	35.8	40.8	45.7	-4.95	PIM
HP	t50	light	11	47543161	20.85-34.1	2.36	11.1	62.6	56	49.3	6.63	PIM
LN	AUC	light	11	48283991	23.4-30.7	7.03	21.2	32.8	41.4	49.9	-8.41	PIM
LN	t10	light	11	48283991	21.9-27.8	6.06	18.7	49.4	42.6	35.7	6.87	PIM
LN	t50	light	11	48283991	21.9-31.7	6.02	17	66.1	56.9	47.7	8.63	PIM
LN	Gmax	light	11	48586795	19.2-30.1	3.64	13.7	1.64	1.81	1.98	-0.17	PIM
HP	t10	light	11	49292479	14.3-35.1	2.55	12.3	43.6	38.4	33.2	4.98	PIM
LN	u8416	light	11	49292479	25.4-32.7	5.35	18.7	30.4	25.3	20.3	5.02	PIM
LN	AUC	light	12	6238531	43.9-58.1	2.93	8	36.17	41.4	46.7	-5.23	PIM
LN	t50	light	12	6238531	45.9-54.8	4.23	11.4	63.7	56.7	49.7	6.97	PIM
LN	u8416	light	12	6238531	43.9-53.8	2.55	8.3	28.7	25.3	22	3.39	PIM
LN	t10	light	12	47269280	51.7-58.1	4.05	11.8	48.3	42.7	37.2	5.56	PIM

The table description is the same as the description for Table 1 with the addition of μ S. lyc, estimated mean of trait associated with MM allele; μ S. lyc, idem for PIM allele; μ het, idem for the heterozygous allele.

The QTL analysis revealed numerous genetic region assumed to be responsible for seed germination qualities as the response to the light condition under two maternal environment. As the matter of fact, the power of QTL detection depends on various factors, for instance, is the heritability of the trait (Foolad et al. 1997). In case of heritability, most of the traits studied in this experiment are considered highly heritable, indicated by the high percentage of this parameter (see Table 9, column heritability). Hence, it can be suggested that most of the differences in seed quality phenotypes were due to genetic variation among individuals in the RIL population. In relation to QTL analysis, high heritability increases the power of QTL detection. Nevertheless, even though the trait has low heritability, the QTL still could be detected if it has a major effect. For instance, four QTL with relatively large effect was detected for the low heritable trait of Gmax in light condition under HP maternal environment (see Table 9). In contrary, there could be more QTLs affecting seed germination qualities in light condition but not detected due its to small effect in this experiment.

The QTL study also reveals the complexity of genetic regulation of seed quality phenotypes as the response of light. Few numbers of QTLs were associated with a single trait while the majority were associated with more than one traits. Among the latter category, some QTLs were detected only in a certain maternal environment. For instance, on chromosome 4 (support interval: 41—55 cM), the QTLs associated with t10 and t50 were detected only in LN but not in HP maternal environment. On the other hand, there were also QLTs that were associated with a certain trait in the different maternal environment. The example for this is QTL for t10, t50, and AUC in chromosome 11 (support interval: 20—30 cM) that were present in both maternal environments. More confusingly, a QTL also discovered to be

associated with the different trait in the different maternal environment. This QTL is found on chromosome 5 (support interval: 0—10 cM), where in LN it is associated with t10 while in HP it was associated with u8416.

The collocation of QTL could be explained by either pleiotropic effect of the gene or close physical linkage of the non-pleiotropic gene. In order to confirm this, a sophisticated statistical test needs to be performed, which is not done in this study (Jiang et al. 1995). In case of QTL in chromosome 11 (support interval: 20—30 cM), it could be assumed that the underlying gene for this QTL has a pleiotropic effect because of the association with many traits, including dark and light germination in both studied maternal environment. The support interval of this QTL is, in fact, overlap with another seed germination traits under stress condition according to the study using similar RIL population (Khan et al. 2012). Furthermore, the presence of the collocating QTL under different conditions support the hypothesis regarding the involvement of common physiological mechanism of seed germination under various environmental condition (Foolad et al. 2007). The utilization of this QTL in marker-assisted breeding would have a great implication since many aspects of seed germination qualities will be enhanced at once under different conditions.

4.3. Candidate genes for light germination was identified by gene co-expression network analysis

The identification of candidate gene for a certain trait by using only QTL data is impractical due to a wide area of genome encompassing the support interval. Therefore, the QTL data was combined with genome-wide expression and gene co-expression data to perform a network analysis and identify the most promising causal genes in the resulting network. In this study, the network analysis was conducted for Gmax and AUC trait and resulted in four candidate genes: Solyc04g051720, Solyc05g014090, Solyc05g042180, and Solyc06g069710. Next, the function of these genes was determined to find their connection to seed germination as the response to light. Due to lack of information of tomato genes, the orthologues of these tomato genes in Arabidopsis were sought and used to predict the function of the candidate genes. Lastly, the expression of the candidate genes was analyzed to validate their involvement in the corresponding trait by qPCR assay.

From functional description retrieved from the publicly available database, all of the orthologues of a candidate or its co-expressed genes identified in this study has connections in some extent to seed germination mechanism or light response. For instance, Solyc04g051720 has a homology to At1g16850,

an unknown transmembrane protein that serves as a binding target of HY5 (Lee et al. 2007; Kleine et al. 2007). HY5 itself is a transcription factor that promotes the expression of light-inducible genes to regulate photomorphogenic development (Osterlund et al. 2000). Therefore, this information suggested that Solyc04g051720, as the orthologous of the HY5-binding target, may play roles in photomorphogenic development. To support this argument, the co-expression of tomato orthologous of Hy5—Solyc08g061130—with Solyc04g051720 was analyzed in TomExpress but the result showed that these two genes are not co-expressed together. Hence, the regulation of Solyc04g051720 may be different with its orthologous in Arabidopsis.

On the other hand, Solyc04g051720 is also co-expressed with Solyc08g078190—an orthologue to *ERF6* (AT4G17490)—and Solyc05g032660—an orthologous to *NOL* (AT5G04900). The former gene is an ethylene transcription factors that is rapidly upregulated as a response to light (Moore et al. 2014; König et al. 2017). Meanwhile, *NOL* is a chlorophyll b reductase gene, which involves in preventing leaf photodamage due to high-light intensity (Sato et al. 2015) and leaf senescence in darkness condition (Zhang et al. 2015). This gene is also expressed during embryogenesis to maintain seed maturation and storability (Nakajima et al. 2012). All thing considered, this information related to the co-expressed gene and their association to light response provide further suggestion related to the involvement of Solyc04g051720 in light germination.

In addition to gene description from the database, the involvement of Solyc04g051720 in light germination was further analyzed by qPCR assay (Figure 7). The result showed that the gene only expressed in PIM, particularly in the dried seed. The QTL for this gene had a positive effect, indicating that the beneficial allele for this gene is coming from MM, which lacks the expression of Solyc04g051720. Therefore, we could assume that the absence of Solyc04g051720 expression has a positive effect on Gmax or, in another word, Solyc04g051720 may play a role as a seed germination inhibitor. The QTL, however, was only detected in light germination of HP maternal environment, which is strange since there is no clear distinction in the expression pattern between lighting condition and maternal environment, at least until 12 hai.

On the other hand, the database search on Solyc05g014090 and Solyc05g042180 provide a little evidence whether these genes are involved in light germination of tomato seed. These two genes are located on chromosome 5 and chosen as the possible candidate gene due to their co-expression with many trans-eQTLs in the gene co-expression network (see Figure 5). In Sol Genomics Network, Solyc05g014090 was described as a histone-binding protein and Solyc05g042180 as an RNA-binding protein suggesting their

role in post-translational and post-transcriptional modification in cellular signaling, respectively. These signaling mechanisms are in fact known to occur during seed germination (Narsai et al. 2017; Nakabayashi et al. 2005). However, the role of this gene in light germination is not known yet. In Arabidopsis, Solyc05g042180 is homologous to AT4G26000 while Solyc05g014090 to AT2G19540. Yet, both genes do not have any connection to seed germination or response to light. However, there is one compelling information that connects Solyc05g042180 with light response in seed germination. It turned out that this gene was co-expressed with Solyc05g025510 that have an Arabidopsis orthologue to *AtWDR26* (*AT5G08560*). *AtWDR26* is a WD-40 repeat protein which expression is induced by light. Furthermore, the over-expression of this gene in light condition reduced the seed germination rate, suggesting the inhibitory nature of *AtWDR26* (Chuang et al. 2015).

As for qPCR assay, only the expression of Solyc05g014090 could be analyzed because the primer for the other gene cannot be optimized. The gene expression analysis of Solyc05g014090 showed that this gene was differentially expressed between two parental lines but only in LN (Figure 8), which may explain the detection of this gene exclusively in such maternal environment. In HP, the transcript of this gene was abundant in the dried seed of PIM and MM. The Solyc05g014090 transcript was then decreasing after 6 and 12 hours of imbibition. In LN, the same pattern was also observed in MM but not in PIM because the expression was already low in the dried seed. Hence, this could be an indication that the regulation of Solyc05g014090, just like Solyc04g051720, occurs before the initial phase of imbibition and later expressed during the very early seed germination phase, leading to the decrease in seed germination capacity. In LN, this was not happening in seed carrying PIM allele, thus the Gmax was not affected. To conclude, Solyc05g014090 may serve as an inhibitor of Gmax and the allele of this gene in PIM was repressed possibly due to its interaction with LN maternal environment.

Last but not the least, there is Solyc06g069710, which is the most promising candidate gene for light germination in this study. In Sol Genomic Network, this gene is described as non-apical meristem (NAM) protein known to be involved in plant development (Ooka et al. 2003). Solyc06g069710 has many orthologous in Arabidopsis. Among them is *ORE1*, which has the most sequence similarity with Solyc06g069710 according to BLAST result. Numerous literature has described the role of this transcription factor as the master regulator of leaf senescence (Qiu et al. 2015; Matallana-Ramirez et al. 2013; Rauf et al. 2013). Furthermore, a study by Balazadeh et al. (2010) showed that ORE1 is also involved in blocking of seed germination under the high saline condition, providing evidence for Solyc06g069710 as the candidate gene in seed germination.

In addition to high saline condition, the expression of ORE1 is induced by the dark condition as well (Balazadeh et al. 2010). Coincidently, the result of qPCR assay in this study also suggests that the expression of its tomato homolog--Solyc06g069710—was upregulated in dark condition (Figure 9). It could be assumed that—because of the expression of Solyc06g069710 was higher in dark—the seed germination phenotype of dark-treated seed was inhibited considering the negative role of this gene in seed germination. This pattern, however, only occurred in MM because, in PIM, no difference was observed in gene expression between dark- and light-treated seed. On the other hand, the upregulation of *ORE1* in dark condition is controlled by AtPIF5 that bind directly to the *ORE1* promoter in Arabidopsis. This interaction might also occur between Solyc06g069710 and the orthologous of *AtPIF5* in tomato. Therefore, the orthologous of *AtPIF5* in tomato was sought in PANTHER – Gene List Analysis to analyze its expression. It turns out that *AtPIF5* and *AtPIF4* genes were originated from a duplication event in Brassicaceae family, explaining the presence of single *SlPIF4* (Solyc07g043580) in *S. lycopersicum* (Rosado et al. 2016). The co-expression *SlPIF4* and Solyc06g069710 were then analyzed in TomExpress and the result showed that these genes were not co-expressed one to another. Hence, the regulation of Solyc06g069710 and Arabidopsis ORE1 gene in dark condition may be different.

4.4. The study limitation

This study has successfully provided a simple and straightforward method for determining causal genes for complex traits using a limited amount of information. However, the identification of a genetic component of complex traits using genetical-genomics approach is not without limitation. The main disadvantage of this approach is that the method only identifies genetic component of the phenotypic variation solely based on the gene expression differences. As dictated by central dogma of molecular biology, the flow of genetic information was described as two-step processes: transcription and translation. Hence, one may assume that once a gene is transcripted into mRNA, it is always translated into a functional protein, leading to the expression of a certain phenotype. This, however, is not always the case. As the matter of fact, there is another mechanism that is not covered in the central dogma that influences phenotypic variation, such as post-translational modification, post-transcriptional modification, and epigenetic modifications. These mechanisms are in fact occurs during seed development and germination (Kawakatsu et al. 2017; Martinez-Andujar et al. 2011; Narsai et al. 2017). which explain that not all gene co-expression network constructed in this study resulted in a promising candidate gene in addition to the other factors. This problem, nevertheless, could be solved by

incorporating data from another level of regulation, such as protein and metabolic QTL (Ligterink et al. 2012).

Another thing that worth considering is the source of genome-wide expression data. Instead of imbibed seed, I used the expression data from dried seed with the hope that the allele expression difference between two parents at this stage take part in determining the seed germination phenotypes at the later stages. In fact, the occurrence of stored RNA in the mature seed is common among plant species (Dure & Waters 2016; Almoguera & Jordano 1992; Ishibashi et al. 1990). Little is known about the function of stored RNA but it is assumed to play a role in protein synthesis during the early phase of seed germination (Nakabayashi et al. 2005). Therefore, there is a possibility that the candidate genes for light germination are differentially expressed in the dried seed of two parental lines and it determines the seed germination performance as the response to light.

4.5. The expression of *PHYA*, *PHYB2*, and *GIGANTEA* in germinating tomato seed

In tomato seed germination, light stimulus for seed germination is perceived by either *PHYA* or *PHYB2*. Particularly and related to this study, *PHYA* is a photoreceptor for a continuous light response while *PHYB2* for short-light exposure (Appenroth et al. 2006) and dark response (Bertram et al. 2004). To study the regulation of these genes in seed germination, an expression analysis was performed using qPCR. The results showed the gene expression of *PHYA* was higher in dark-treated compared to light-treated seed for all genotype under LN and HP maternal environment. The downregulation of *PHYA* gene expression in this study is consistent with the result of the same gene expression in Arabidopsis (Somers & Quail 1995) and tobacco (Adam et al. 1994). This phenomenon may be caused by autoregulation of PHYA protein to its gene transcript. In oat, this mechanism has been discovered where light-induced PHYA-fr formation triggers a certain second messenger cascade, which in turn lead to rapid repression of *PHYA* transcript (Quail 1994).

As for *PHYB2,* the expression in MM was higher in dark-treated seed and no difference between two lighting conditions in PIM. To the best of my knowledge, no study has described the autoregulation of PHYB2 so far. One thing for sure is that the gene regulation was slightly decreasing after imbibition and the rate was lower in MM under dark condition. Therefore, the fact that the gene was downregulated suggest the occurrence of autoregulation in PHYB2 since this gene is a photoreceptor for both short light pulse and dark response. It could be assumed as well that the rate of PHYB2 transcription inhibition in

MM in dark condition is not as rapid as in light condition. Furthermore, the more stable expression of *PHYB2* in MM under dark condition could be an explanation of its higher germination onset (t10) and rate (t50) in this condition in comparison to light condition.

Beside *PHYA* and *PHYB2*, another gene presumed to be involved in continuous light response is GIGANTEA (*GI*). The result showed that at 12 hai there were no differences in the expression of this gene between light and dark imbibition in all genotypes under two maternal environment. It is worth mentioning that in MM the gene expression of light-treated seed is higher than dark-treated seed at 6 hai. Furthermore, there was also no distinctive fluctuation of gene expression between dried seed until 12-hours imbibed seed, except in PIM of LN where the expression is higher in dried seed but decreased after imbibition. Hence, the gene expression analysis in this experiment did not support the involvement of *GIGANTEA* in light germination.

GIGANTEA is known to promote seed germination in Arabidopsis by regulating PHYA-mediated very-low-fluence response (Oliverio et al. 2007). In tomato, the expression pattern of *GIGANTEA* in germinating seed has been studied previously by Rutitzky et al. (2009) and Auge et al. (2009). However, the results of this gene expression appeared to be contradicting. Rutitzky et al. (2009) showed that the expression of *GIGANTEA* at was increasing to promote seed germination when PHYA-mediated HIR was initiated through hourly Rp. In contrary, Auge et al. (2009) indicate that GI was upregulated and inhibit seed germination through the same response. However, these two results cannot be compared offhandedly because the cultivar, RNA measurement method, light treatment, and magnitude of seed germination used in the experiments were different.

5. CONCLUSION AND RECOMMENDATION

Based on the result of this experiment, I conclude that:

1. Light influence t10 and t50 of *S. lycopersicum* cv. Moneymaker and *S. pimpinellifolium*. However, the response of both species was different. In *S. lycopersicum* cv. Moneymaker, light increased the t10 and t50 while in *S. pimpinellifolium* these parameters were decreased.
2. Twenty-nine QTLs were found for light germination. Most of the positive allele of these QTLs were originated from *S. pimpinellifolium*.
3. Network analysis was performed using the combination of QTL mapping, QTL analysis, and gene co-expression data, resulting in four candidate genes for light germination traits. These candidate genes are Solyc04g051720, Solyc05g014090, Solyc05g042180, and Solyc06g069710. Among these, Solyc06g069710 is the most promising candidate gene based on expression analysis and database search of its orthologous.
4. The expression *PHYB2* were higher in the dark-treated seed of *S. lycopersicum* whereas no difference between lighting conditions was observed in *S. pimpinellifolium*. As for *PHYA*, the expression was higher in the dark-treated seed of all genotype. Meanwhile, the expression of *GIGANTEA* between light and dark-treated seed were the same.

Several recommendations were considered for the next research on this topic. First, specific light wavelength and exposure duration can be used to specify the response of seed germination. Second, another gene co-expression network can be constructed for another QTLs discovered in this study. Furthermore, the incorporation of functional annotation to each gene in the network may be useful to gene prioritization. Third, the candidate genes in this network should be further validated for its involvement in light germination by reverse genetics study.

REFERENCE

Adam, E. et al., 1994. The developmental and tissue-specific expression of tobacco phytochrome-A genes. *The Plant Journal*, 6(3), pp.283–293.

Almoguera, C. & Jordano, J., 1992. Developmental and environmental concurrent expression of sunflower dry-seed-stored low-molecular-weight heat-shock protein and Lea mRNAs. *Plant Molecular Biology*, 19(5), pp.781–792.

Appenroth, K.J. et al., 2006. Tomato seed germination: Regulation of different response modes by phytochrome B2 and phytochrome A. *Plant, Cell and Environment*, 29(4), pp.701–709.

Aud, F.F. & Ferraz, I.D.K., 2012. Seed size influence on germination responses to light and temperature of seven pioneer tree species from the Central Amazon. *Anais da Academia Brasileira de Ciências*, 84(3), pp.759–766. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22886162.

Auge, G.A. et al., 2009. Gene expression analysis of light-modulated germination in tomato seeds. *New Phytologist*, 183(2), pp.301–314.

Balazadeh, S. et al., 2010. A gene regulatory network controlled by the NAC transcription factor ANAC092/AtNAC2/ORE1 during salt-promoted senescence. *Plant Journal*, 62(2), pp.250–264.

Baskin, C.C. & Baskin, J.M., 2014. *Ecologically Meaningful Germination Studies*, Available at: http://linkinghub.elsevier.com/retrieve/pii/B9780124166776000020.

Bertram, L., Diara, C. & Lercari, B., 2004. Seed germination in tomato is controlled by physiological interactions of multiple phytochromes. *Advances in Horticultural Science*, 18(1), pp.3–10. Available at: http://www.jstor.org/stable/42882280.

Bewley, J.D. et al., 2013. *Seeds*, New York, NY: Springer New York. Available at: http://link.springer.com/10.1007/978-1-4614-4693-4.

Borthwick, H.A. et al., 1954. Action of Light on Lettuce-Seed Germination. *Botanical Gazette*, 115(3), pp.205–225. Available at: https://doi.org/10.1086/335817.

Borthwick, H., Hendricks, S. & Parker, M., 1952. A reversible photoreaction controlling seed germination. *Proceedings of the National Academy of Sciences*, 38(8), pp.662–666. Available at: http://www.pnas.org/content/38/8/662.full.pdf.

Casal, J.J., Sanchez, R.A. & Botto, J.F., 1998. Modes of action of phytochromes. *Journal of Experimental Botany*, 49(319), pp.127–138. Available at: http://jxb.oxfordjournals.org/lookup/doi/10.1093/jxb/49.319.127.

Chuang, H. wen et al., 2015. An Arabidopsis WDR protein coordinates cellular networks involved in light, stress response and hormone signals. *Plant Science*, 241, pp.23–31. Available at: http://dx.doi.org/10.1016/j.plantsci.2015.09.024.

Collard, B.C.Y. et al., 2005. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica*, 142(1–2), pp.169–196.

Dekkers, B.J.W. et al., 2012. Identification of reference genes for RT-qPCR expression analysis in arabidopsis and tomato seeds. *Plant and Cell Physiology*, 53(1), pp.28–37.

DeVicente, M.C. & Tanksley, S.D., 1993. QTL analysis of transgressive segregation in an interspecific tomato cross. *Genetics*, 134(2), pp.585–596.

Doganlar, S., Frary, A. & Tanksley, S.D., 2000. The genetic basis of seed-weight variation: Tomato as a model system. *Theoretical and Applied Genetics*, 100(8), pp.1267–1273.

Druka, A. et al., 2010. Expression quantitative trait loci analysis in plants. *Plant Biotechnology Journal*, 8(1), pp.10–27.

Dure, L. & Waters, L., 2016. Long-Lived Messenger RNA : Evidence from Cotton Seed Germination Published by : American Association for the Advancement of Science Linked references are available on JSTOR for this article : , 147(3656), pp.410–412.

Eckstein, A. et al., 2016. Mobilization of storage materials during light-induced germination of tomato (Solanum lycopersicum) seeds. *Plant Physiology and Biochemistry*, 105, pp.271–281.

Fleige, S. & Pfaffl, M.W., 2006. RNA integrity and the effect on the real-time qRT-PCR performance. *Molecular Aspects of Medicine*, 27(2–3), pp.126–139.

Flores, J., González-Salvatierra, C. & Jurado, E., 2016. Effect of light on seed germination and seedling shape of succulent species from Mexico. *Journal of Plant Ecology*, 9(2), pp.174–179.

Foolad, M.R. et al., 1997. Mapping QTLs conferring salt tolerance during germination in tomato by selective genotyping. *Molecular Breeding*, 3(4), pp.269–277. Available at: http://www.springerlink.com/openurl.asp?id=doi:10.1023/A:1009668325331%5Cnfile:///Users/paillecy/Dropbox/Papers2/Articles/1997/Foolad/Molecular Breeding 1997 Foolad.pdf publication/doi/10.1023/A:1009668325331.

Foolad, M.R., Subbiah, P. & Zhang, L., 2007. Common QTL affect the rate of tomato seed germination under different stress and nonstress conditions. *International Journal of Plant Genomics*, 2007.

Gharoobi, B., 2006. Effects of seed size on seedlings characteristics of five barley cultivars. , pp.265–270.

Groot, S.P.C. & Karssen, C.M., 1987. Gibberellins regulate seed germination in tomato by endosperm weakening: a study with gibberellin-deficient mutants. *Planta*, 171(4), pp.525–531.

Gui, M., Wang, X. & Huang, W., 1991. The effect of light and exogenous gibbereilic acid on respiration pathways durii ^ germination of tomato seeds. , pp.403–407.

Hayashi, E., Aoyama, N. & Still, D.W., 2008. Quantitative trait loci associated with lettuce seed germination under different temperature and light environments. *Genome / National Research Council Canada = Génome / Conseil national de recherches Canada*, 51(11), pp.928–47. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18956026.

Ishibashi, N., Yamauchi, D. & Minamikawa, T., 1990. Vigna unguiculata. , pp.59–64.

Jiang, C. et al., 1995. Multiple Trait Analysis of Genetic Mapping for Quantitative Trait Loci. *Genetics*, 140(3), pp.1111–1127.

Jiménez-Gómez, J.M., Wallace, A.D. & Maloof, J.N., 2010. Network analysis identifies ELF3 as a QTL for the shade avoidance response in arabidopsis. *PLoS Genetics*, 6(9).

Joosen, R.V.L. et al., 2009. Advances in genetical genomics of plants. *Current genomics*, 10(8), pp.540–549.

Joosen, R.V.L. et al., 2010. Germinator: A software package for high-throughput scoring and curve fitting of Arabidopsis seed germination. *Plant Journal*, 62(1), pp.148–159.

Kawakatsu, T. et al., 2017. Dynamic DNA methylation reconfiguration during seed development and germination. *Genome Biology*, 18(1), pp.1–12.

Khan, N. et al., 2012. Exploring the Natural Variation for Seedling Traits and Their Link with Seed Dimensions in Tomato. *PLoS ONE*, 7(8), pp.20–31.

Kleine, T. et al., 2007. Genome-Wide Gene Expression Analysis Reveals a Critical Role for CRYPTOCHROME1 in the Response of Arabidopsis to High Irradiance. *Plant Physiology*, 144(3), pp.1391–1406. Available at: http://www.plantphysiol.org/cgi/doi/10.1104/pp.107.098293.

König, K. et al., 2017. The significance of glutathione and ascorbate in modulating the retrograde high light response in *Arabidopsis thaliana* leaves. *Physiologia Plantarum*. Available at: http://doi.wiley.com/10.1111/ppl.12644.

Kubista, M. et al., 2006. The real-time polymerase chain reaction. *Molecular Aspects of Medicine*, 27(2–3), pp.95–125.

Lee, J. et al., 2007. Analysis of Transcription Factor HY5 Genomic Binding Sites Revealed Its Hierarchical Role in Light Regulation of Development. *the Plant Cell Online*, 19(3), pp.731–749. Available at:

http://www.plantcell.org/cgi/doi/10.1105/tpc.106.047688.

Lercari, B. & Lipucci di Paola, M., 1991. Photoregulation of seed germination of wild-type and of an aurea-mutant of tomato. *Physiologia Plantarum*, 83(2), pp.265–268.

Ligterink, W., Joosen, R.V.L. & Hilhorst, H.W.M., 2012. Unravelling the complex trait of seed quality: using natural variation through a combination of physiology, genetics and -omics technologies. *Seed Science Research*, 22(S1), pp.S45–S52.

Livak, K.J. & Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and. *Methods*, 25, pp.402–408.

Mancinelli, A.L., Borthwick, H.A. & Hendricks, S.B., 1966. Phytochrome Action in Tomato-Seed Germination. *Botanical Gazette*, 127(1), pp.1–5.

Martinez-Andujar, C. et al., 2011. Post-transcriptional gene regulation during seed germination and stand establishment. *Acta horticulturae*.

Martínez-Andújar, C., Martin, R.C. & Nonogaki, H., 2012. Seed traits and genes important for translational biology-highlights from recent discoveries. *Plant and Cell Physiology*, 53(1), pp.5–15.

Matallana-Ramirez, L.P. et al., 2013. NAC transcription factor ORE1 and Senescence-Induced BIFUNCTIONAL NUCLEASE1 (BFN1) constitute a regulatory cascade in arabidopsis. *Molecular Plant*, 6(5), pp.1438–1452.

Moore, M., Vogel, M.O. & Dietz, K.J., 2014. The acclimation response to high light is initiated within seconds as indicated by upregulation of AP2/ERF transcription factor network in Arabidopsis thaliana. *Plant Signaling and Behavior*, 9(10), pp.1–4.

Morgante, M. & Salamini, F., 2003. From plant genomics to breeding practice. *Current Opinion in Biotechnology*, 14(2), pp.214–219.

Murali, K.S., 1997. Patterns of seed size, germination and seed viability of tropical tree species in southern India. *Biotropica*, 29(3), pp.271–279.

Nakabayashi, K. et al., 2005. Genome-wide profiling of stored mRNA in Arabidopsis thaliana seed germination: Epigenetic and genetic regulation of transcription in seed. *Plant Journal*, 41(5), pp.697–709.

Nakajima, S. et al., 2012. Chlorophyll b Reductase Plays an Essential Role in Maturation and Storability of Arabidopsis Seeds. *Plant Physiology*, 160(1), pp.261–273. Available at: http://www.plantphysiol.org/cgi/doi/10.1104/pp.112.196881.

Narsai, R. et al., 2017. Extensive transcriptomic and epigenomic remodelling occurs during Arabidopsis thaliana germination. *Genome Biology*, 18(1), pp.1–18.

Nomaguchi, M., Nonogaki, H. & Morohashi, Y., 1995. Development of galactomannan-hydrolyzing activity in the micropylar endosperm tip of tomato seed prior to germination. *Physiologia Plantarum*, 94(1), pp.105–109.

Oliverio, K.A. et al., 2007. GIGANTEA regulates phytochrome A-mediated photomorphogenesis independently of its role in the circadian clock. *Plant physiology*, 144(1), pp.495–502. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1913770&tool=pmcentrez&rendertype=abstract.

Van Ooijen, J.W., 2009. MapQTL 6, Software for the mapping of quantitative trait loci in experimental populations of dipoid species. *Kyazma B.V., Wageningen, Netherlands*, (April).

Ooka, H. et al., 2003. Comprehensive analysis of NAC family genes in *Oryza sativa* and *Arabidopsis thaliana*. *DNA Research*, 10(January), pp.239–247. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15029955.

Osterlund, M.T. et al., 2000. Targeted destabilization of HY5 during light-regulated development of Arabidopsis. *Nature*, 405(6785), pp.462–466.

Paran, I. & Zamir, D., 2003. Quantitative traits in plants: Beyond the QTL. *Trends in Genetics*, 19(6), pp.303–306.

Pearson, T.R.H. et al., 2002. GERMINATION ECOLOGY OF NEOTROPICAL PIONEERS: INTERACTING EFFECTS OF ENVIRONMENTAL CONDITIONS AND SEED SIZE. , 1(3), pp.316–325. Available at: http://www.jstor.org/stable/3072016.

Pet, G. & Garretsen, F., 1983. Genetical and environmental factors influencing seed size of tomato (Lycopersicon esculentum Mill.) and the effect of seed size on growth and development of tomato. *Euphytica*, 32, pp.711–718. Available at: http://www.springerlink.com/index/x64537q7585p2j78.pdf.

Porter, N.G. & Clark, S.M., 1979. Effect of temperature and light on the germination of seed of solanum aviculare and s. laciniatum. *New Zealand Journal of Experimental Agriculture*, 7(3), pp.307–310.

Qiu, K. et al., 2015. EIN3 and ORE1 Accelerate Degreening during Ethylene-Mediated Leaf Senescence by Directly Activating Chlorophyll Catabolic Genes in Arabidopsis. *PLoS Genetics*, 11(7), pp.1–20.

Quail, P., 1994. Phytochrome genes and their expression. *Photomorphogenesis in Plants*, pp.71–104.

Rauf, M. et al., 2013. ORE1 balances leaf senescence against maintenance by antagonizing G2-like-mediated transcription. *EMBO Reports*, 14(4), pp.382–388.

Rosado, D. et al., 2016. Phytochrome Interacting Factors (PIFs) in Solanum lycopersicum: Diversity, evolutionary history and expression profiling during different developmental processes. *PLoS ONE*,

11(11), pp.1–21.

Rutitzky, M. et al., 2009. Comparative genomic analysis of light-regulated transcripts in the Solanaceae. *BMC genomics*, 10, p.60. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2644711&tool=pmcentrez&rendertype=abstract.

Sato, R., Ito, H. & Tanaka, A., 2015. Chlorophyll b degradation by chlorophyll b reductase under high-light conditions. *Photosynthesis Research*, 126(2–3), pp.249–259.

Shannon, P. et al., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, (13), pp.2498–2504.

Shichijo, C. et al., 2001. Phytochrome A-mediated inhibition of seed germination in tomato. *Planta*, 213(5), pp.764–769.

SHROPSHIRE W., J.R., KLEIN, W.H. & ELSTAD, V.B., 1961a. ACTION SPECTRA OF PHOTOMORPHOGENIC INDUCTION AND PHOTOINACTIVATION OF GERMINATION IN ARABIDOPSIS THALIANA. *Plant and Cell Physiology*, 2(1), pp.63–69. Available at: http://dx.doi.org/10.1093/oxfordjournals.pcp.a077664.

SHROPSHIRE W., J.R., KLEIN, W.H. & ELSTAD, V.B., 1961b. ACTION SPECTRA OF PHOTOMORPHOGENIC INDUCTION AND PHOTOINACTIVATION OF GERMINATION IN ARABIDOPSIS THALIANA. *Plant and Cell Physiology*, 2(1), pp.63–69.

Somers, D.E. & Quail, P.H., 1995. Temporal and spatial expression patterns of PHYA and PHYB genes in Arabidopsis. *The Plant journal : for cell and molecular biology*, 7(3), pp.413–427.

Souza, M.L. & Fagundes, M., 2014. Seed Size as Key Factor in Germination and Seedling Development of Copaifera langsdorffii (Fabaceae). *American Journal of Plant Sciences*, (July), pp.2566–2573.

Taiz, L. & Zeiger, E., 2010. *Plant Physiology*,

Tarquis, A.M. & Bradford, K., 1992. *Prehydration and Priming Treatments that Advance Germination also Increase the Rate of Deterioration of Lettuce Seeds*,

Thompson, P.A., 1974. Characterisation of the germination responses to temperature of vegetable seeds. I. Tomatoes. *Scientia Horticulturae*, 2(1), pp.35–54.

van Tuinen, A. et al., 1995. Far-red light-insensitive, phytochrome A-deficient mutants of tomato. *MGG Molecular & General Genetics*, 246(2), pp.133–141.

WHITTINGTON, W.J. et al., 1965. Analysis of Variation in the Rates of Germination and Early Seedling Growth in Tomato. *Annals of Botany*, 29(1), pp.59–71. Available at: +.

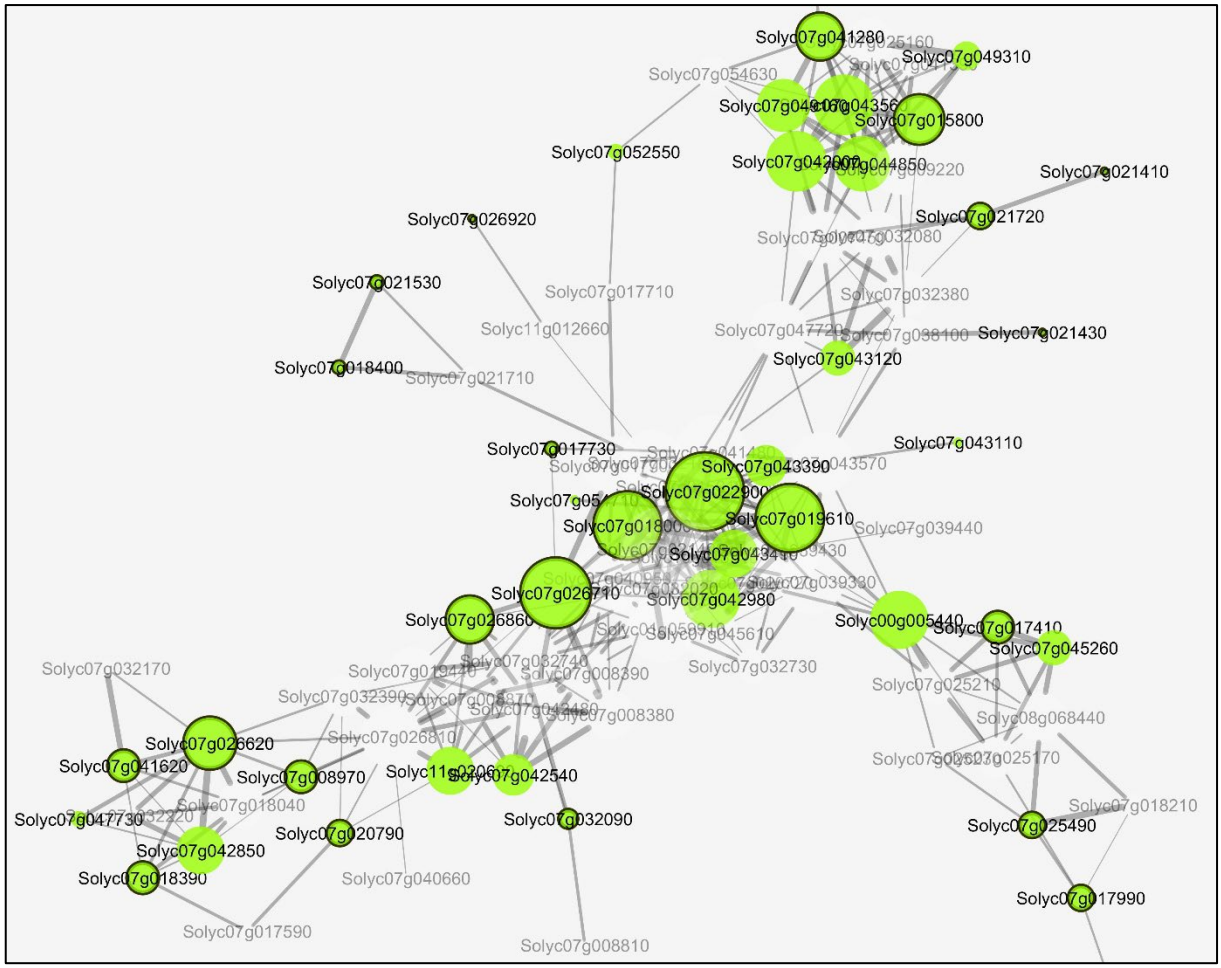
Xia, Q., Ando, M. & Seiwa, K., 2016. Interaction of seed size with light quality and temperature regimes as germination cues in 10 temperate pioneer tree species. *Functional Ecology*, 30(6), pp.866–874.

Zhang, Y. et al., 2015. PHYTOCHROME-INTERACTING FACTOR 5 (PIF5) positively regulates dark-induced senescence and chlorophyll degradation in Arabidopsis. *Plant Science*, 237, pp.57–68.

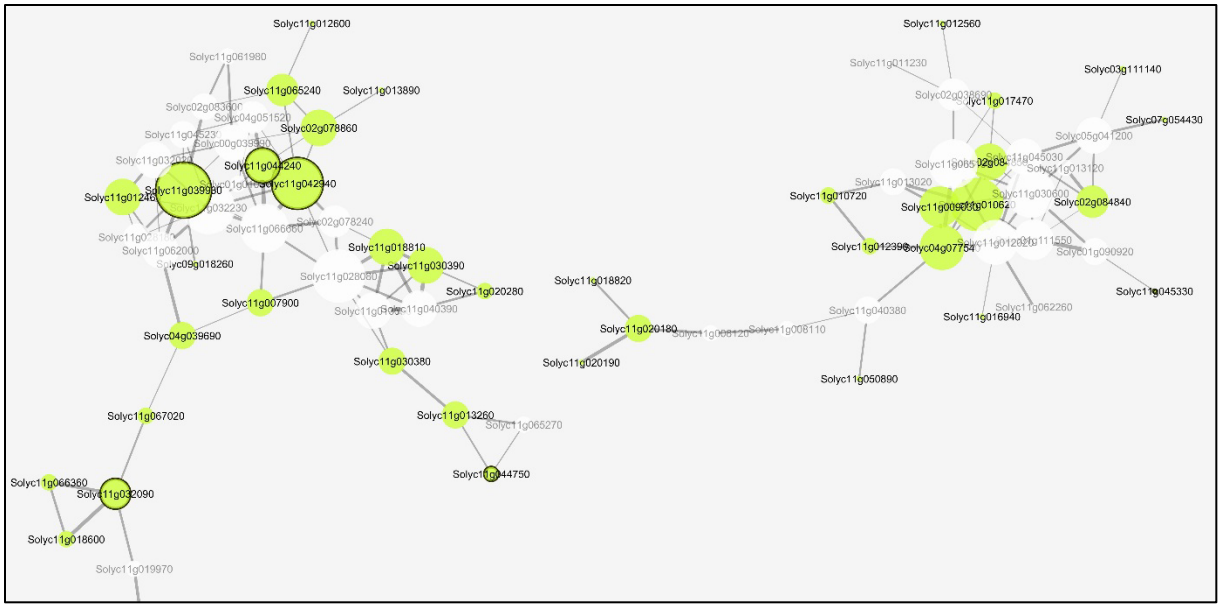
Zhou, J., Deckard, E.L. & Messersmith, C.G., 2005. Factors Affecting Eastern Black Nightshade (Solanum ptycanthum) Seed Germination. *Weed Science*, 53(5), pp.651–656. Available at: http://www.jstor.org/stable/4047033.

Zouine, M. et al., 2017. TomExpress, a unified tomato RNA-Seq platform for visualization of expression data, clustering and correlation networks. *The Plant Journal*, pp.727–735.

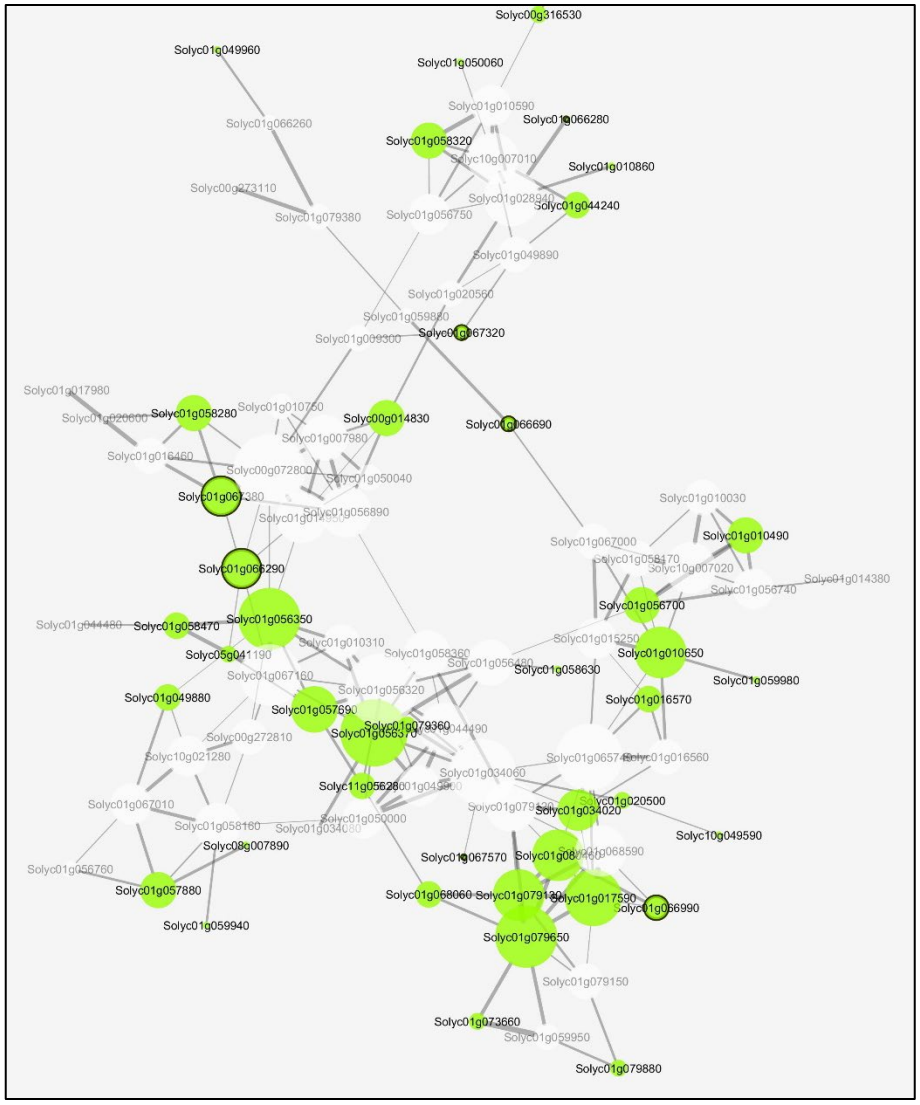
APPENDIX



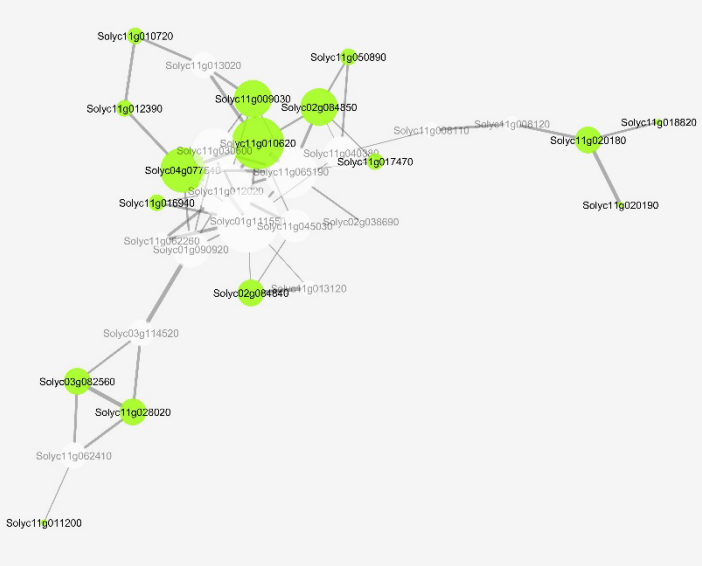
Appendix 1. Network analysis for 102 genes located in the QTL in chromosome 7 for Gmax in HP maternal environment (peak locus: 28,076,245 mb). Genes are represented by nodes. Nodes with borders are genes located in the support interval of the QTL in chromosome 7. Larger nodes has more degree of connection compared to smaller nodes. Moreover, nodes with green color represent genes that has higher expression (or lower if it is an inhibitor) in dried PIM seed under HP maternal environment compared to the other genotype and maternal environment. These genes are connected to each other by edges. Thick edges indicate strong correlation and the other way around for thin edges. Note that several insignificant nodes and edges are not shown in the figure to maintain the image resolution.



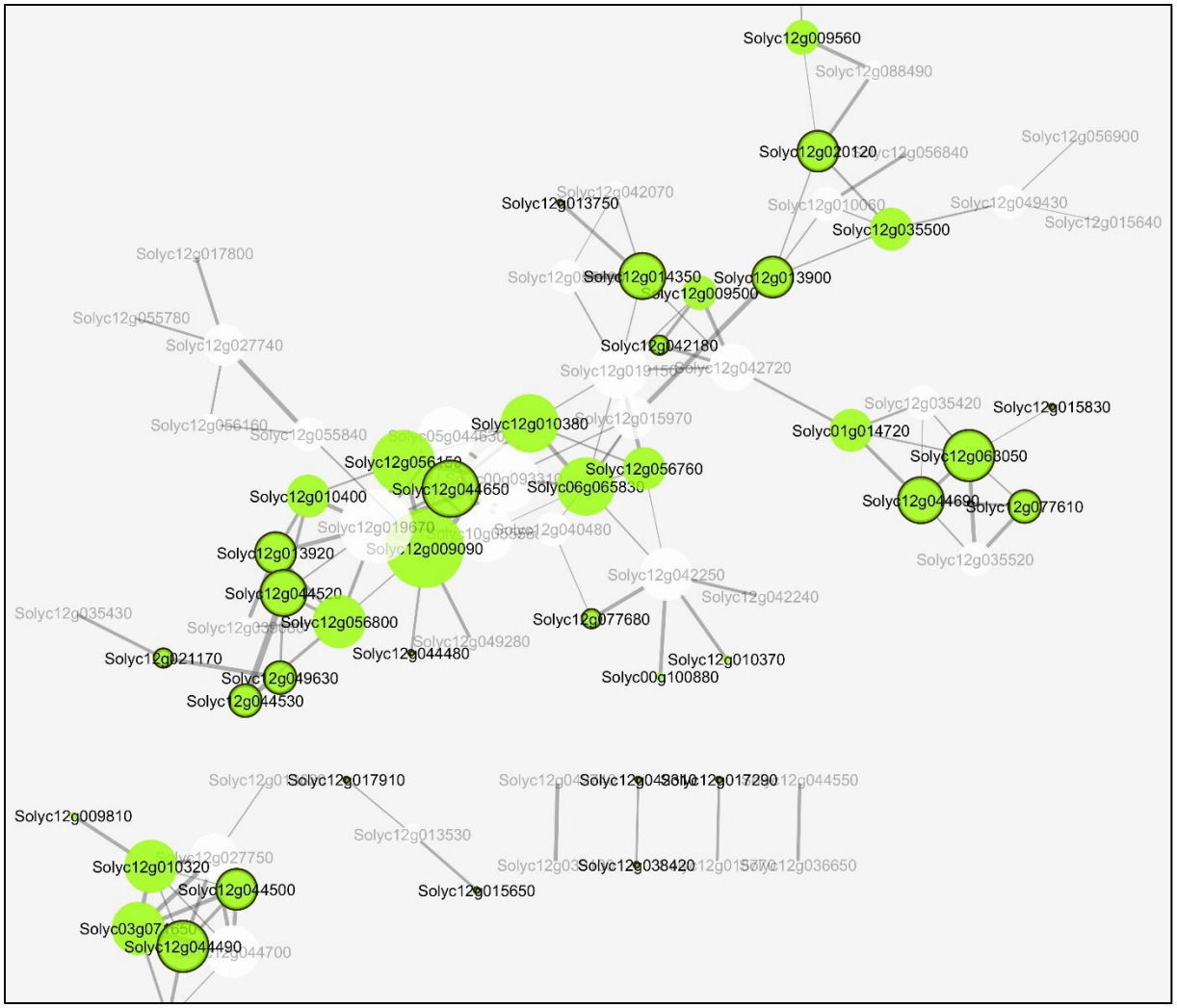
Appendix 2. Network analysis for 93 genes located in the QTL in chromosome 11 for Gmax in LN maternal environment (peak locus: 48,586,795 mb). Genes are represented by nodes. Nodes with borders are genes located in the support interval of the QTL in chromosome 11. Larger nodes has more degree of connection compared to smaller nodes. Moreover, nodes with green color represent genes that has higher expression (or lower if it is an inhibitor) in dried PIM seed under LN maternal environment compared to the other genotype and maternal environment. These genes are connected to each other by edges. Thick edges indicate strong correlation and the other way around for thin edges. Note that several insignificant nodes and edges are not shown in the figure to maintain the image resolution.



Appendix 3. Network analysis for 93 genes located in the QTL in chromosome 1 for AUC in LN maternal environment (peak locus: 76,672,459 mb). Genes are represented by nodes. Nodes with borders are genes located in the support interval of the QTL in chromosome 1. Larger nodes has more degree of connection compared to smaller nodes. Moreover, nodes with green color represent genes that has higher expression (or lower if it is an inhibitor) in dried PIM seed under LN maternal environment compared to the other genotype and maternal environment. These genes are connected to each other by edges. Thick edges indicate strong correlation and the other way around for thin edges. Note that several insignificant nodes and edges are not shown in the figure to maintain the image resolution.

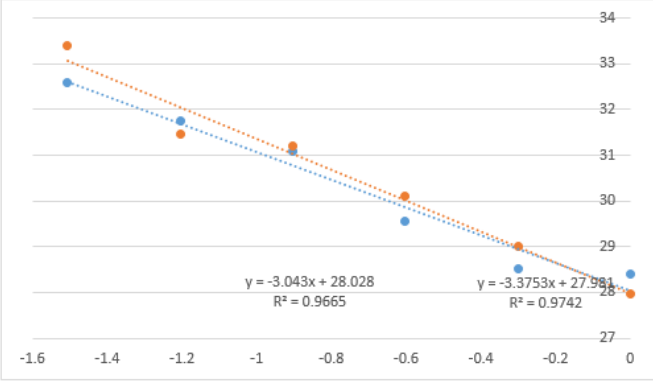


Appendix 4. Network analysis for 89 genes located in the QTL in chromosome 11 for AUC in LN maternal environment (peak locus: 48,283,771 mb). Genes are represented by nodes. Nodes with borders are genes located in the support interval of the QTL in chromosome 11. Larger nodes has more degree of connection compared to smaller nodes. Moreover, nodes with green color represent genes that has higher expression (or lower if it is an inhibitor) in dried PIM seed under LN maternal environment compared to the other genotype and maternal environment. These genes are connected to each other by edges. Thick edges indicate strong correlation and the other way around for thin edges. Note that several insignificant nodes and edges are not shown in the figure to maintain the image resolution.



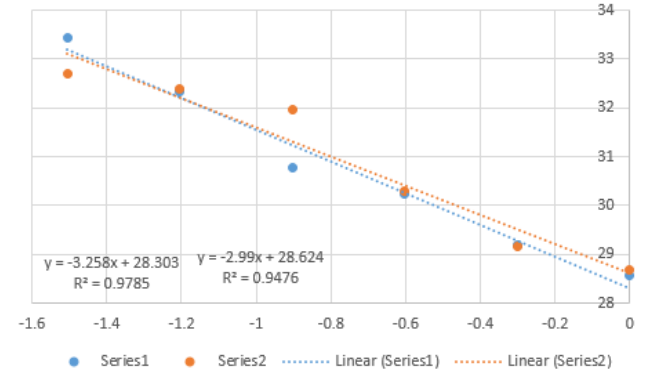
Appendix 5. Network analysis for 81 genes located in the QTL in chromosome 12 for AUC in LN maternal environment (peak locus: 6,238,531mb). Genes are represented by nodes. Nodes with borders are genes located in the support interval of the QTL in chromosome 12. Larger nodes has more degree of connection compared to smaller nodes. Moreover, nodes with green color represent genes that has higher expression (or lower if it is an inhibitor) in dried PIM seed under LN maternal environment compared to the other genotype and maternal environment. These genes are connected to each other by edges. Thick edges indicate strong correlation and the other way around for thin edges. Note that several insignificant nodes and edges are not shown in the figure to maintain the image resolution.

Primer pair	cDNA dillution	dill. Fact.	log (1/dill.fact)	ct value sample 1	ct value sample 2	average		slope	efficiency	R2
51720	200	0	0	28.40290312	27.96275835	28.18283074	1	-3.043	113.1	0.9665
	400	2	-0.30103	28.53173687	29.01245797	28.77209742	2	-3.3753	97.8	0.9742
	800	4	-0.60206	29.5549966	30.10449604	29.82974632	**avg**	**-3.20915**	**104.9**	**0.9704**
	1600	8	-0.90309	31.0953169	31.19683001	31.14607345				
	3200	16	-1.20412	31.73965846	31.46558778	31.60262312				
	6400	32	-1.50515	32.58232163	33.38488271	32.98360217				



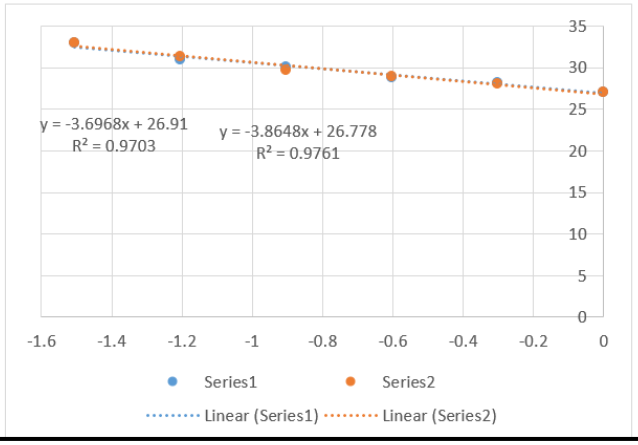
Appendix 5. Primer efficiency and linear regression for Solyc04g051720 using serial dilution

Primer pair	cDNA dillution	dill. Fact.	log (1/dill.fact)	ct value sample 1	ct value sample 2	average		slope	efficiency	R2
									%	
69710	200	0	0	28.56201863	28.68956752	28.62579308	1	-3.25	103.1	0.9785
	400	2	-0.30103	29.19706064	29.16443994	29.18075029	2	-2.99	116.0	0.9476
	800	4	-0.60206	30.23266314	30.29730353	30.26498334	**avg**	**-3.12**	**109.2**	**0.9631**
	1600	8	-0.90309	30.76601223	31.97910816	31.3725602				
	3200	16	-1.20412	32.33272901	32.40472717	32.36872809				
	6400	32	-1.50515	33.43916417	32.7095884	33.07437629				



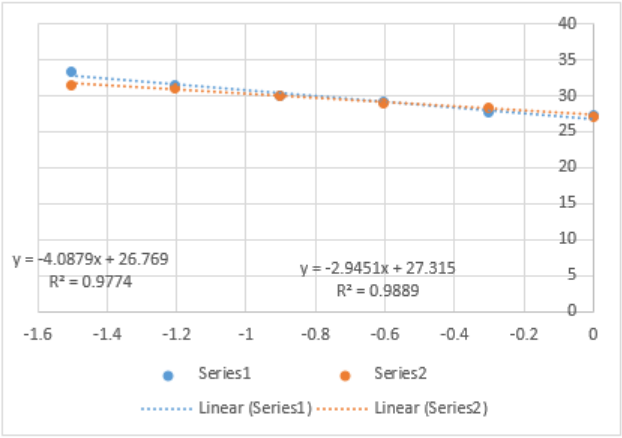
Appendix 6. Primer efficiency and linear regression for Solyc06g069710 using serial dilution

Primer pair	cDNA dillution	dill. Fact.	log (1/dill.fact)	ct value sample 1	ct value sample 2	average		slope	efficiency	R2
									%	
14090	200	0	0	27.07	27	27.035	**1**	-3.697	86.4	0.9703
	400	2	-0.30103	28.22	28.09	28.155	**2**	-3.865	81.4	0.9761
	800	4	-0.60206	28.87	28.92	28.895	**avg**	**-3.781**	**83.9**	**0.9732**
	1600	8	-0.90309	30.08	29.73	29.905				
	3200	16	-1.20412	30.9	31.36	31.13				
	6400	32	-1.50515	33.01	33.02	33.015				



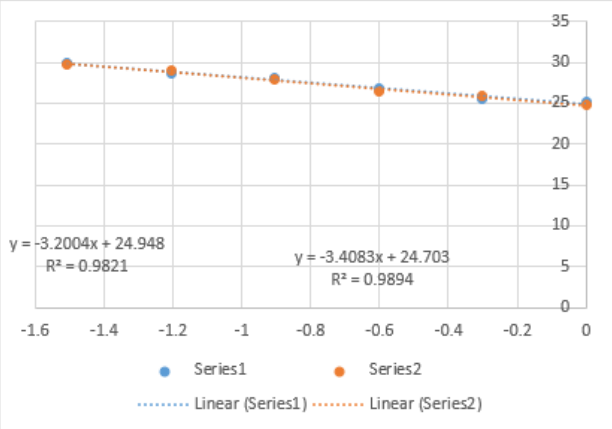
Appendix 7. Primer efficiency and linear regression for Solyc04g014090 using serial dilution

Primer pair	cDNA dillution	dill. Fact.	log (1/dill.fact)	ct value sample 1	ct value sample 2	average		slope	efficiency	R2
									%	
GI	200	0	0	27.19	27.1	27.145	**1**	-4.0879	75.6	0.9785
	400	2	-0.30103	27.75	28.42	28.085	**2**	-2.9451	118.5	0.9476
	800	4	-0.60206	29.18	29.06	29.12	**avg**	**-3.5165**	**92.5**	**0.9631**
	1600	8	-0.90309	30.05	30.05	30.05				
	3200	16	-1.20412	31.55	31	31.275				
	6400	32	-1.50515	33.35	31.56	32.455				



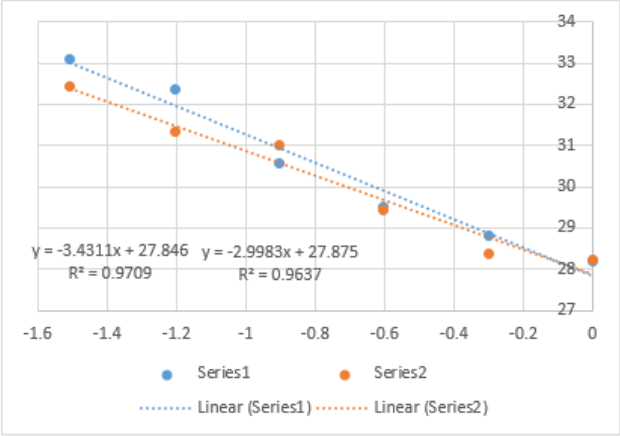
Appendix 8. Primer efficiency and linear regression for *GIGANTEA* using serial dilution

Primer pair	cDNA dillution	dill. Fact.	log (1/dill.fact)	ct value sample 1	ct value sample 2	average		slope	efficiency	R2
									%	
PhyA	200	0	0	25.18	24.76	24.97	1	-3.2004	105.3	0.9821
	400	2	-0.30103	25.59	25.83	25.71	2	-3.4083	96.5	0.9894
	800	4	-0.60206	26.81	26.39	26.6	avg	**-3.30435**	**100.7**	**0.9858**
	1600	8	-0.90309	28.11	27.9	28.005				
	3200	16	-1.20412	28.58	28.98	28.78				
	6400	32	-1.50515	29.87	29.75	29.81				



Appendix 9. Primer efficiency and linear regression for *PHYA* using serial dilution

Primer pair	cDNA dillution	dill. Fact.	log (1/dill.fact)	ct value sample 1	ct value sample 2	average		slope	efficiency	R2
									%	
PhyB	200	0	0	28.19	28.21	28.2	1	-3.4311	95.6	0.9709
	400	2	-0.30103	28.83	28.36	28.595	2	-2.9983	115.5	0.9637
	800	4	-0.60206	29.52	29.45	29.485	avg	**-3.2147**	**104.7**	**0.9673**
	1600	8	-0.90309	30.58	31	30.79				
	3200	16	-1.20412	32.36	31.34	31.85				
	6400	32	-1.50515	33.09	32.43	32.76				



Appendix 10. Primer efficiency and linear regression for *PHYB* using serial dilution