

RESEARCH ARTICLE

A Shortest-Path-Based Method for the Analysis and Prediction of Fruit-Related Genes in *Arabidopsis thaliana*

Liucun Zhu¹ , Yu-Hang Zhang² , Fangchu Su¹, Lei Chen³, Tao Huang^{2*}, Yu-Dong Cai^{1*}

1 School of Life Sciences, Shanghai University, Shanghai, People's Republic of China, **2** Institute of Health Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, People's Republic of China, **3** College of Information Engineering, Shanghai Maritime University, Shanghai, People's Republic of China

 These authors contributed equally to this work.

* tohuangtao@126.com (TH); cai_yud@126.com (YDC)



CrossMark
click for updates

 OPEN ACCESS

Citation: Zhu L, Zhang Y-H, Su F, Chen L, Huang T, Cai Y-D (2016) A Shortest-Path-Based Method for the Analysis and Prediction of Fruit-Related Genes in *Arabidopsis thaliana*. PLoS ONE 11(7): e0159519. doi:10.1371/journal.pone.0159519

Editor: Quan Zou, Tianjin University, CHINA

Received: May 23, 2016

Accepted: July 5, 2016

Published: July 19, 2016

Copyright: © 2016 Zhu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This study was supported by the National Natural Science Foundation of China (31371335 and 31501045), Shanghai Sailing Program and The Youth Innovation Promotion Association of Chinese Academy of Sciences (CAS) (2016245). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Biologically, fruits are defined as seed-bearing reproductive structures in angiosperms that develop from the ovary. The fertilization, development and maturation of fruits are crucial for plant reproduction and are precisely regulated by intrinsic genetic regulatory factors. In this study, we used *Arabidopsis thaliana* as a model organism and attempted to identify novel genes related to fruit-associated biological processes. Specifically, using validated genes, we applied a shortest-path-based method to identify several novel genes in a large network constructed using the protein-protein interactions observed in *Arabidopsis thaliana*. The described analyses indicate that several of the discovered genes are associated with fruit fertilization, development and maturation in *Arabidopsis thaliana*.

Introduction

In botany, fruit usually refers to the seed-bearing structure in angiosperms and results from the maturation of flowers, the gynoecium of which contributes to the formation of all fruit structures. The whole development and maturation processes of fruits are precisely regulated by functional genes and can be artificially divided into three main steps: fertilization, development and maturation. During the fertilization process, activation of the gibberellin signaling pathways contributes to fruit growth [1], and genes such as DELLA, ARFs and Aux/IAA have been reported to participate in the gibberellin signaling pathways and to contribute to both maintenance of the hormonal balance and the regulation of cell division and size alteration, functions that are crucial in the fertilization process [2, 3]. In addition to the fertilization process, the development stage of fruit also involves various functional genes. Fruits can be divided into fleshy and dry fruits, and the development of these different types of fruits involves different pathways and genes. For example, the development of tomato, as a fleshy fruit, involves cell division and expansion of the ovary tissues, and several genes contribute to these processes, including Fw2.2, which is responsible for approximately 30% of the variation in tomato size

and contributes to the regulation of the size of fleshy fruits [4]. In addition, genes such as SUN, OVATE and FASCIATED in tomato are associated with the shape of the tomato fruit during its development [5–7]. Thus, the fertilization and development processes of fruits (including tomato) are precisely regulated. The maturation process of dry fruits can be further separated into two stages, namely maturation and dehiscence, whereas that of fleshy fruits can be separated into maturation and ripening. During the maturation and dehiscence of dry fruits, genes such as SHP1 and SHP2 are positively regulated by the IND gene, which further contributes to the maturation process. All three genes (SHP1, SHP2 and IND) are further regulated by a core gene for fruit maturation, FRUITFULL (FUL). The fertilization, development and maturation processes of fruits are all regulated by specific genes, implying the need to identify functional regulatory genes that contribute to the fruit-associated pathways activated during the fertilization, development and maturation processes of fruits.

Arabidopsis thaliana (also known as thale cress, mouse-ear cress or Arabidopsis) is a small cruciferous plant that originated from Eurasia [8]. *Arabidopsis thaliana*, which was first named *Pilosella siliquosa*, was first discovered in 1577 in the Harz Mountains in Germany by Johannes Thal, a German botanist [9]. During the 300 years following its discovery, *Arabidopsis thaliana* was identified and confirmed to be native to Europe, Asia, and northwestern Africa [8]. In 1943, Friedrich Laibach first proposed *Arabidopsis thaliana* as a model organism. Since then, the application of *Arabidopsis thaliana* in biological sciences, particularly the fields of mutation and genomics, has gradually increased and become popularized [10, 11]. *Arabidopsis thaliana* presents four main advantages regarding its use as a classical and functional model in the biological sciences, particularly genetics. First, *Arabidopsis thaliana* is short (7–40 cm), which enables scientists to grow the plant on a large scale in a small amount of space. Second, *Arabidopsis thaliana* has a strong reproductive ability [12], which permits the generation of a large number of plant seedlings within a short time and reduces the cost of experiments. Additionally, as *Arabidopsis thaliana* is a self-pollinated plant, its fertilization process is less affected by the external environment than those of cross-pollinated plants, making it easier to control extraneous variables during experiments [13]. Most of the genes of wild-type *Arabidopsis thaliana* are highly homozygous, making it easy to induce specific mutations and to precisely construct the corresponding mutants via various physical and chemical methods [14]. Furthermore, as a cruciferous plant, *Arabidopsis thaliana* has a genome of only five chromosomes and approximately 100 million base pairs, which greatly decreases the difficulty and cost of whole-genome sequencing (WGS) [15]. In addition, the WGS of *Arabidopsis thaliana* was completed in 2000 by the Arabidopsis Genome Initiative, popularizing the application of *Arabidopsis thaliana* as an experimental model and the study of functional genes of Arabidopsis during the whole lifecycle of *Arabidopsis thaliana* [16].

In our daily life, we always pay attention to the fruits of plants, such as apples and pears. In botany, a fruit is defined as the specialized seed-bearing structure designed to protect the seeds [17]. The fruit of *Arabidopsis thaliana* is a silique or seedpod that develops from the fertilized gynoecium and dehisces at maturity, allowing the seeds contained inside to spread. Because *Arabidopsis thaliana* is a self-pollinated plant, intrinsic genetic regulatory factors play an irreplaceable role during the fruit development and maturity processes of *Arabidopsis thaliana*. For example, mutations in two functional genes, *KANADI1* and *KANADI2*, directly contribute to the development of the lateral organs (a crucial part of the fruit of *Arabidopsis thaliana*) [18]. Additionally, *CRABS CLAW* (*CRC*) plays a specific role in the gynoecium, the developmental source of the fruit, validating the crucial role of genetic factors in various aspects of fruit development in *Arabidopsis thaliana* [19]. The differentiation of tissues necessary for the opening and dehiscence of the fruit has been considered the core biological process during fruit development [20]. Genes such as *FUL* contribute to the leaf development and meristem

identity processes, the phenotype of which has been widely reported in the fruits of *Arabidopsis thaliana* [21]. As we have mentioned above, the genetic contributions to the biological processes of *Arabidopsis thaliana*, especially those of the fruit, have been widely studied and validated, indicating the significance of the identification of the core genes of the physiological processes in *Arabidopsis thaliana*.

Recently, several studies adopted the shortest path algorithm to investigate some important biological and medical problems, such as identifying the genetic determinants in yeast longevity study [22], discovery of novel disease genes [23–28], prediction of novel carcinogenic chemicals [29], etc. In this study, using *Arabidopsis thaliana* as a research model as well as validated functional genes, we propose a shortest-path-based method for the prediction of novel genes that contribute to the biological processes related to fruit. A group of genes were accessed by the shortest-path-based method, and several genes were confirmed to contribute to fruit development and maturity through a review of recent publications. The new findings obtained in this study will further promote investigations of the physiological mechanisms underlying fruit development and maturation.

Materials and Methods

Materials

Validated genes. Validated fruit-associated genes in *Arabidopsis thaliana* were extracted from Plant Ontology (PO, <http://www.plantontology.org/download>) [30]. First, we downloaded the plant_ontology.obo file from PO (accessed on March 24, 2015), which contained the PO term structure. All of the terms related to fruit and their children terms were considered fruit PO terms, which included fruit (PO:0009001) and its parts (PO:0004707, PO:0008001, PO:0004536, PO:0004535, PO:0008002, PO:0025268, PO:0000033, PO:0008003, PO:0009087, and PO:0009084). This analysis revealed 994 *Arabidopsis thaliana* genes annotated with these PO terms, which are provided in [S1 Table](#) and are considered fruit-related genes. The set consisting of these 994 genes was denoted S_v .

PPIs of *Arabidopsis thaliana*. PPIs play important roles in executing, modulating and maintaining the activities and functions of organisms. Proteins that can interact with each other share common functions [23–26, 31–34]. According to the validated genes mentioned in Section “Validated genes”, we can search for novel fruit-related genes using the PPIs of *Arabidopsis thaliana*. In this study, we adopted the PPIs reported in STRING (<http://string-db.org/>, version 9.1) [35], which were collected from the following sources: genomic context, high-throughput experiments, (conserved) coexpression, and previous knowledge. The PPIs indicate not only the direct (physical) associations but also the indirect (functional) associations between proteins. To extract the PPIs of *Arabidopsis thaliana*, we downloaded a file called “[protein.links.v9.1.txt.gz](#)” from STRING, which contains the PPIs of several organisms, and extracted the lines starting with “3702.”. A total of 3,123,482 PPIs of *Arabidopsis thaliana*, which included 25,123 proteins, were obtained. Each of these PPIs contains two proteins, which are represented by their Ensembl IDs, and one score ranging from 150 to 999. This score measures the strength of the interaction: an interaction with a high score indicates that the corresponding proteins have a high probability of interacting with each other. We denoted the score of the interaction between proteins p_1 and p_2 as $Q(p_1, p_2)$.

Network construction

The PPIs of *Arabidopsis thaliana* were used to construct a large network: the 25,123 proteins were defined as the nodes of the network, and the 3,123,482 PPIs were the edges. To further indicate the strength of the interactions in the network, an edge weight was assigned to each

edge. The maximum value of the interaction score was 999, and according to the shortest-path algorithm, a low weight suggests a strong association between the endpoints of the corresponding edge. Thus, the weight of edge e was defined as follows:

$$w(e) = 1000 - Q(p_1, p_2) \quad (1)$$

where p_1 and p_2 are the proteins represented by the two endpoints of edge e . The information of each edge in the network, including its end-points and its weight, is provided in [S2](#), [S3](#) and [S4](#) Tables.

Shortest-path-based method

A shortest-path algorithm is a classic graph algorithm that has been widely used in several fields. For example, shortest-path algorithms have been adopted for the discovery of novel disease genes [23–28]. In this study, a shortest-path algorithm was used to discover novel fruit-related genes based on the validated genes mentioned in Section “Validated genes”. The basic theory is that proteins that can interact with each other have common functions [23–26, 31–34]. If one further considers the interaction scores, this fact can be generalized to state that the proteins in a protein-protein interaction with a high score are more likely to share common functions than those in an interaction with a low score. For the network constructed as described in Section “Network construction”, this statement can be converted into the following: nodes (proteins) that are endpoints of an edge assigned a low weight are more likely to share common functions than those that are endpoints of an edge assigned a high weight. Furthermore, if there are several nodes, denoted n_1, n_2, \dots, n_k such that n_i and n_{i+1} are the endpoints of an edge with a low weight, all of the nodes may share common functions. These nodes may all lie on the shortest path connecting nodes n_1 and n_k , and all nodes other than n_1 and n_k may share the functions shared by nodes n_1 and n_k . Thus, we searched all shortest paths connecting any two of the validated genes and extracted the nodes included in these paths. Because the endpoints of these paths represent genes in *Arabidopsis thaliana* that are related to the biological processes of fruit, the genes corresponding to the extracted nodes may also be related to these biological processes. The validated genes may also be found as inner nodes in these paths, and these genes were retained to indicate the utility of the shortest-path-based method. The extracted genes that were not validated genes were deemed candidate fruit-related genes in *Arabidopsis thaliana*. In addition, we counted the number of paths containing each extracted gene and termed this value as betweenness, which can indicate the associations between the extracted genes and fruit development in *Arabidopsis thaliana*.

Based on the validated genes, we can identify novel genes in the constructed network. The topological structure of the network is a potential factor that may influence the results because some genes in *Arabidopsis thaliana* can exhibit general associations with other genes, suggesting that the corresponding nodes in the network are general hubs. These genes are more easily identified by the shortest-path algorithm but have less or no association with fruit development in *Arabidopsis thaliana* and should thus be excluded. As a result, a permutation test was performed to exclude these genes. A total of 1,000 gene sets, each with the same size of S_v , were randomly produced. For each set, the shortest paths connecting any two genes in the set were searched in the network to determine the betweenness of the genes extracted as described in the above paragraph. Accordingly, 1,000 betweenness values were obtained for the 1,000 randomly produced sets, and one betweenness value was found for S_v . The permutation false discovery rate (permutation FDR) was then calculated for each obtained gene as follows:

$$FDR(g) = \frac{M}{1000} \quad (2)$$

where M is the number of betweenness values of randomly produced sets that are greater than that found for S_v . A gene with a high permutation FDR exhibits general associations with the 1,000 randomly produced sets and is not specific to the validated gene set S_v . Thus, we set 0.05 as the threshold for the permutation FDR to exclude these genes, i.e., genes with permutation FDRs less than 0.05 were selected as the candidate genes.

Results and Discussion

In this study, the shortest-path-based method was used to identify novel fruit-related genes in *Arabidopsis thaliana*. The procedure and some of the key results are illustrated in Fig 1. This section provides a detailed description of the results and analyses.

Results of the shortest-path-based method

As mentioned in Section “Shortest-path-based method”, all shortest paths connecting any two genes in S_v were searched in the network. The genes corresponding to the inner nodes in these paths were extracted, and their betweenness was determined. As a result, 3,375 genes were

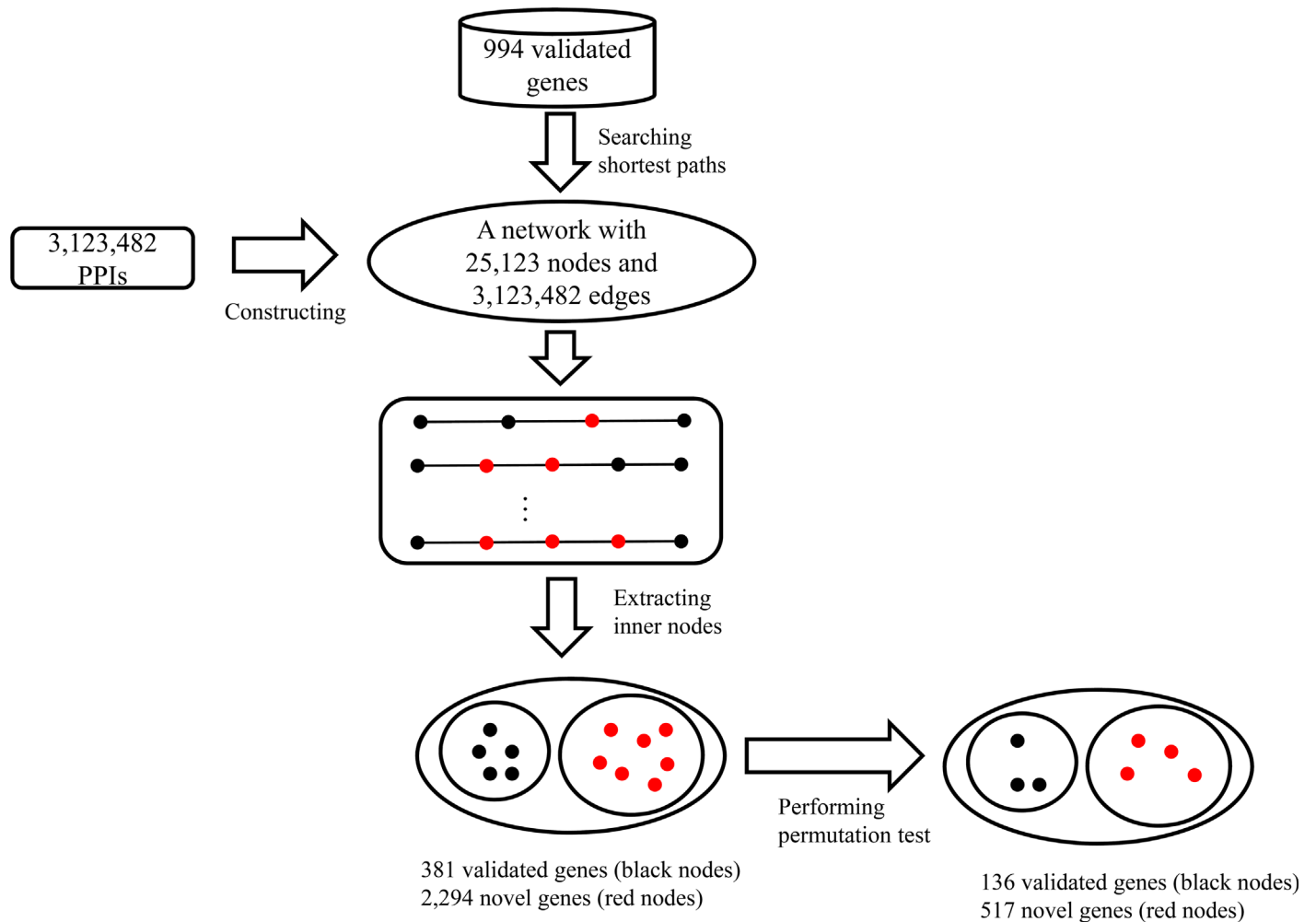


Fig 1. Procedures of and results obtained using the shortest-path-based method for the identification of novel fruit-related genes in *Arabidopsis thaliana*.

doi:10.1371/journal.pone.0159519.g001

obtained; of these, 381 genes were validated genes, and 2,294 genes were novel genes. Detailed information for these genes is provided in [S5 Table](#).

To control the general genes among the 3,375 genes, a permutation test was performed. We thus calculated the permutation FDR for each gene (provided in [S5 Table](#)). After excluding genes with permutation FDRs greater than or equal to 0.05, we identified 653 genes, including 136 validated genes and 517 novel genes, all of which are provided in [S6 Table](#). The 517 novel genes were deemed to exhibit associations with the biological processes of the fruit of *Arabidopsis thaliana* and were considered candidate genes.

Analysis of the validated and candidate genes

The fruit of *Arabidopsis thaliana* has been widely studied, and the complex regulatory mechanisms underlying its development and maturity have also been extensively studied. In this study, based on the shortest-path-based method, several novel genes were identified and deemed to contribute to the development and maturity of the fruit of *Arabidopsis thaliana*. Furthermore, some validated genes were identified by our method, indicating the robustness of the shortest-path-based method. This section provides a detailed analysis of some representative genes.

Analysis of the validated genes. The shortest-path-based method identified 136 validated genes. The five genes with the highest betweenness values were singled out and are listed in [Table 1](#). These five genes are also extensively discussed in the following section.

AT5G63310 (encoding Arabidopsis nucleoside diphosphate kinase 2) is a functional gene that contributes to the response to oxidative stress and UV and is involved in phytochrome-mediated light signaling [36–38]. It has been confirmed that UV, as an extrinsic factor, may interfere with the normal development of fruit in *Arabidopsis thaliana* [39]. Because this gene has been found to be expressed in the flower and fruit of *Arabidopsis thaliana*, the responsive and protective function of the gene may contribute to the fruit-associated biological processes in *Arabidopsis thaliana*. The gene ontology analysis of known genes revealed that known fruit-related genes are enriched in specific exogenous stimulus response-associated biological processes, such as **GO: 0006970** (response to osmotic stress) and **GO: 0009416** (response to light stimulus), indicating that the prediction of AT5G63310 is not a special case.

Another gene, **AT1G75950** (encoding Arabidopsis Skp1 homologue 1) is also on the list of known fruit-associated genes, encoding the core component of the E3 ubiquitin ligase and contributes to the auxin-activated signaling pathway [40, 41]. As part of the SCF (Skp1-CUL1-F-box protein) complexes, this gene contributes to embryogenesis and seedling development, which is significant for the early development of *Arabidopsis thaliana* [42]. Another predicted

Table 1. Five validated genes identified by the shortest-path-based method.

Gene	Also known as	Protein	Function	Reference
AT5G63310	ATNDPK2	Arabidopsis nucleoside diphosphate kinase 2	GTP biosynthetic process, response to UV, response to hydrogen peroxide	[38, 39, 75]
AT1G75950	ATSKP1	Arabidopsis Skp1 homologue 1	Core component of the SCF family of E3 ubiquitin ligases, embryogenesis and seedling development	[40, 42]
AT5G57360	ZTL	Clock-associated PAS protein ZTL	SCF-dependent proteasomal ubiquitin-dependent protein catabolic process, circadian rhythm, entrainment of the circadian clock by photoperiod, flower development	[43, 44]
AT1G23190	PGM	Phosphoglucomutase 3	Carbohydrate metabolic process, glucose metabolic process	[45, 46]
AT4G33010	GLDP1	Glycine decarboxylase P-protein 1	Oxidation reduction, glycine catabolic process, response to cadmium ion	[47–49],

doi:10.1371/journal.pone.0159519.t001

gene, **AT5G57360**, also contributes to the SCF-associated biological processes [43]. Due to its interaction with both CRY1 and phyB, such gene contributes to the regulation of circadian rhythms during the development of the gynoecium, validating its role in fruit development in *Arabidopsis thaliana* [44]. Known fruit-related genes in *Arabidopsis thaliana* are enriched in certain biological processes, including **GO: 0009416** (response to light stimulus), indicating that more functional genes in the list of known genes, such as the two above-mentioned genes, contribute to the development and maturity processes of fruits.

Another gene, **AT1G23190** [also known as phosphoglucosmutase 3 (PGM3)], is also on our list of known genes. Such gene not only participates in photosynthesis associated biological processes and glucose metabolic processes as widely reported but according to recent publications, contribute to the development of reproductive tissues, including seeds and fruits [45, 46]. **AT4G33010** [encoding glycine decarboxylase P-protein 2 (GLDP2)] in our prediction list contributes to the dehydrogenation processes of glycine in the fruit [47]. This gene has been reported to contribute to a specific biological process, **GO: 0055114** (oxidation reduction), in which our known genes are enriched, validating its core role during fruit development and maturation [48]. Additionally, as a cadmium-associated gene (**GO: 0046686**), AT4G33010 has been confirmed to contribute to the development and maturation processes of the fruits in various biological systems [49].

Analysis of novel genes. In addition to the validated genes, we obtained several novel genes that may contribute to fruit development in *Arabidopsis thaliana*. We then similarly selected the top six novel genes (listed in Table 2) with the highest betweenness values and analyzed these genes, as described in the following paragraphs.

The first predicted gene, **AT1G09570** (encoding far-red elongated hypocotyl 2), contributes to the regulation of photomorphogenesis [50]. Although no direct evidence confirms that AT1G09570 contributes to fruit development and maturity, the biological process it regulates, photomorphogenesis, plays an irreplaceable role during fruit maturation, implying that AT1G09570 may also contribute to fruit associated biological processes, validating our prediction of such light-detection-associated genes [51, 52]. Similarly to AT1G09570, **AT2G18790** (encoding phytochrome B), which contributes to photomorphogenesis, is also on our list of predicted genes [53]. As mentioned above, photomorphogenesis is crucial for fruit development in *Arabidopsis thaliana*, validating the crucial fruit-related role of this gene [51]. Light is one of the most crucial exogenous stimulants for plants. These two predicted genes contribute to the biological processes associated with photo-stimulation, which can be clustered into the specific biological process of response to endogenous stimulus (**GO: 0009719**), corresponding

Table 2. Six candidate genes identified by the shortest-path-based method.

Gene	Also known as	Protein	Function	Reference
AT1G09570	FHY2	Far-red elongated hypocotyl 2	Detection of visible light, gravitropism, negative regulation of translation, photomorphogenesis	[50–52]
AT5G59440	ATTMPK	<i>Arabidopsis thaliana</i> thymidylate kinase 1	dTDP biosynthetic process, dTTP biosynthetic process, dUDP biosynthetic process	[54]
AT2G18790	PHYB	Phytochrome B	Detection of visible light, entrainment of circadian clock, gravitropism, jasmonic acid-mediated signaling pathway, photomorphogenesis	[51, 53]
AT4G02570	ATCUL1	Auxin-resistant 6	SCF complex assembly, auxin-activated signaling pathway, cell cycle, embryo development ending in seed dormancy	[56, 58]
AT5G20570	ATRBX1	Regulator of cullins1	SCF-dependent proteasomal ubiquitin-dependent protein catabolic process	[58–60]
AT4G29040	RPT2A	Regulatory particle AAA-ATPase 2A	Female gamete generation, leaf morphogenesis	[64]

doi:10.1371/journal.pone.0159519.t002

with the GO enrichment analysis of all of the predicted genes. Considering the validated role of photomorphogenesis in fruit development, our newly predicted gene AT1G09570 and AT2G187790 may definitely contribute to fruit associated signaling pathways. In addition to these two genes, another gene, **AT5G59440** (encoding *Arabidopsis thaliana* thymidylate kinase), contributes to the TDP/TTP biosynthesis-associated biological processes, particularly during the fruit-associated fertilization processes [54]. This gene has been widely identified in reproduction-associated tissues of *Arabidopsis thaliana*, including the fruit, validating the relationship between this gene and the development and maturity processes of fruit. Validated in rice, barley, maize and *Arabidopsis thaliana*, thymidylate kinase associated genes have been identified in various reproductive tissues, implying their potential functions for fruit development and maturation [55].

The SCF complex, as we mentioned above, contributes to the specific development and maturation processes of the fruit of *Arabidopsis thaliana* [42]. Included in our list of predicted genes, **AT4G02570** (encoding Auxin-resistant 6) contributes to the SCF complex [56]. Considering the role of SCF complex for fruit development and maturation, such gene may further participates in fruit-related regulatory processes [56]. In addition, this gene contributes to the maturation of the reproductive structures and tissues in *Arabidopsis thaliana*, which is also closely related to fruit-associated biological processes, validating our prediction [57]. Additionally, another novel gene on our list of predicted genes, **AT5G20570** (regulator of cullins 1), contributes to the same biological processes as AT4G02570, indicating their similar contribution to fruit development in *Arabidopsis thaliana* [58]. In addition to its participation in the SCF complex, AT5G20570, which is considered a multi-functional gene, also contributes to CUL4-ROC1-DDB1-PRL1 E3 ligase [59, 60]. Such ligase has been reported to interact with DWD proteins which has been confirmed to contribute to the fruit development and formation of specific resistances in various plants including *Arabidopsis thaliana*, validating the irreplaceable role of our predicted novel gene, AT5G20570 for fruit associated biological processes [61, 62]. Another gene, **AT4G29040** (as known as regulatory particle AAA-ATPase 2A), a DNA methylation associated gene, contributes to root meristem maintenance, gametogenesis, and DNA methylation [63]. This gene has been reported to contribute to fruit development by regulating cell division, proliferation and expansion [64]. The gene induces a weak defect in 26S proteasome activity and leads to enlargement of the fruit (increased cell size and decreased cell number), validating the prediction that our screened gene AT4G29040 contributes to fruit development and maturation [64].

Cross-talk analysis of homological genes in rice and wheat

As we all know, during the evolution, the genome of related organisms (*e.g.* animals, plants, *etc.*) share a fair degree of homology, implying that genes identified in one specie (*e.g.* *Arabidopsis thaliana*) may function similarly in other species like rice or wheat. As we all know, not the *Arabidopsis thaliana* but rice and wheat are two of the major food crops in the world, implying that the identification of reproduction associated genes and pathways may contribute to the genomic modification and further improve the output and quality of such crops. As we have analyzed above, the genes we have identified above may all contribute to reproduction associated pathways (*e.g.* fruit fertilization, development and maturation) in *Arabidopsis thaliana*, suggesting that the homologous genes of such screened genes in rice and wheat may also contribute to fruit associated biological processes and may be one of the core production related regulatory factors. Based on recent publications, we screened out the homologues genes in rice or wheat of our predicted fruit associated genes in *Arabidopsis thaliana* and validated their underlying functions for reproduction. **AT5G63310**, as we have mentioned above, is a

validated fruit associated gene in *Arabidopsis thaliana*. According to the microarray data of wheat germ extract, the homologous gene of AT5G63310 in wheat, **NDPK2** has been confirmed to express in wheat germ extract and may play a similar role with AT5G63310, validating that the homologue of our screened genes may also contribute to fruit associated pathways in other related plants [65]. Besides the homologous gene of such gene (AT5G63310) in rice, **Os12g0548300**, has also been reported to contribute to fruit development and the formation of specific resistance to the hazardous environment, tightly related to the quality and output of such crop [66, 67]. Apart from such validated genes, the homologue genes of our newly predicted genes have also been identified to be candidate fruit associated genes. Take one of our predicted gene, **AT5G59440**, which is also known as *Arabidopsis thaliana* thymidylate kinase 1, as an example. Based on current database for rice genome (e.g. RiceData, RiTE, etc.), we identified the homologous gene of AT5G59440 in rice, named as **Os07g44630** [68]. As a member of thymidylate kinase family, such gene has been confirmed to be identified in the bran of rice fruit and play an irreplaceable role in seed development [69]. As for wheat, the family of thymidylate kinase has been confirmed to contribute to the head sprouting process in wheat, implying its specific role in fruit associated biological processes [70]. Wheat blue dwarf (WBD) is a severe disease for wheat, inducing reproductive abnormality [71, 72]. The homologous gene of our predicted gene **AT5G59440**, **TMK**, as a member of the thymidylate kinase family, has been identified to be dysregulated expressed in the abnormal tissues of such disease, validating its potential role for the fruit associated biological processes in wheat [70].

Just like we have analyzed above, some homologous genes in rice and wheat in our screened out genes in *Arabidopsis thaliana* have been confirmed to contribute to reproductive associated biological processes, according to some recent publications. Such results validates the efficacy and accuracy of our newly presented computational methods and further provide a new research method and perspective into the identification of fruit associated genes in industrial crops.

Conclusions

This study used a shortest-path-based method for the analysis and prediction of fruit-related genes in *Arabidopsis thaliana*. Based on Plant Ontology database, we identified a group of candidate fruit associated genes in *Arabidopsis thaliana*. Considering the protein-protein interactions, several novel genes were identified using our method, and some of these genes were confirmed to be related to fruit development and maturation. The proposed computational method is a useful tool for the identification of novel genes. Further, comparing with some commercial crops like wheat and rice, our screened out fruit associated genes in *Arabidopsis thaliana* were further validated and showed specific inter-species functional conservation. All in all, the newly found genes yielded our computational methods may provide new insights for the investigation of the fruit of *Arabidopsis thaliana* and certain commercial crops. In this study, we only used protein information to investigate fruit-related genes in *Arabidopsis thaliana* because the protein information is much abundant and complete. In future, some other information, such as microRNA [73, 74], will be considered to add into our method to give a more extensive study.

Supporting Information

S1 Table. 994 validated genes related to fruit of *Arabidopsis thaliana*.
(PDF)

S2 Table. Edge information in the constructed network (part one).
(TXT)

S3 Table. Edge information in the constructed network (part two).
(TXT)

S4 Table. Edge information in the constructed network (part three).
(TXT)

S5 Table. Genes identified by the shortest path-based method.
(PDF)

S6 Table. 653 genes filtering by a permutation test, where 136 are validated genes and 517 are novel genes.
(PDF)

Acknowledgments

This study was supported by the National Natural Science Foundation of China (31371335 and 31501045), Shanghai Sailing Program and The Youth Innovation Promotion Association of Chinese Academy of Sciences (CAS) (2016245).

Author Contributions

Conceived and designed the experiments: TH YDC. Performed the experiments: LZ YHZ LC. Analyzed the data: LZ YHZ FS. Contributed reagents/materials/analysis tools: LZ YHZ FS. Wrote the paper: LZ YHZ.

References

1. El-Sharkawy I, Sherif S, El Kayal W, Mahboob A, Abubaker K, Ravindran P, et al. Characterization of gibberellin-signalling elements during plum fruit ontogeny defines the essentiality of gibberellin in fruit development. *Plant Mol Biol*. 2014; 84(4–5):399–413. doi: [10.1007/s11103-013-0139-8](https://doi.org/10.1007/s11103-013-0139-8) PMID: [WOS:000330982900003](https://pubmed.ncbi.nlm.nih.gov/25000003/).
2. Ha S, Tran LS. Understanding plant responses to phosphorus starvation for improvement of plant tolerance to phosphorus deficiency by biotechnological approaches. *Crit Rev Biotechnol*. 2014; 34(1):16–30. doi: [10.3109/07388551.2013.783549](https://doi.org/10.3109/07388551.2013.783549) PMID: [WOS:000331846500002](https://pubmed.ncbi.nlm.nih.gov/25000002/).
3. Aloisi I, Cai G, Serafini-Fracassini D, Del Duca S. Polyamines in Pollen: From Microsporogenesis to Fertilization. *Front Plant Sci*. 2016; 7:155. doi: [10.3389/fpls.2016.00155](https://doi.org/10.3389/fpls.2016.00155) PMID: [26925074](https://pubmed.ncbi.nlm.nih.gov/26925074/); PubMed Central PMCID: PMC4757701.
4. Guo M, Simmons CR. Cell number counts—The fw2.2 and CNR genes and implications for controlling plant fruit and organ size. *Plant Sci*. 2011; 181(1):1–7. doi: [10.1016/j.plantsci.2011.03.010](https://doi.org/10.1016/j.plantsci.2011.03.010) PMID: [WOS:000291512900001](https://pubmed.ncbi.nlm.nih.gov/200291512900001/).
5. Sun XQ, Pang H, Li MM, Chen JQ, Hang YY. Tracing the origin and evolution of plant TIR-encoding genes. *Gene*. 2014; 546(2):408–16. doi: [10.1016/j.gene.2014.04.060](https://doi.org/10.1016/j.gene.2014.04.060) PMID: [WOS:000339150800038](https://pubmed.ncbi.nlm.nih.gov/2500339150800038/).
6. Yu H, Jiang WZ, Liu Q, Zhang H, Piao MX, Chen ZD, et al. Expression Pattern and Subcellular Localization of the Ovate Protein Family in Rice. *Plos One*. 2015; 10(3):e0118966. doi: [10.1371/journal.pone.0118966](https://doi.org/10.1371/journal.pone.0118966) PMID: [WOS:000351275000044](https://pubmed.ncbi.nlm.nih.gov/25000351275000044/).
7. Pautler M, Eveland AL, LaRue T, Yang F, Weeks R, Lunde C, et al. FASCIATED EAR4 Encodes a bZIP Transcription Factor That Regulates Shoot Meristem Size in Maize. *Plant Cell*. 2015; 27(1):104–20. doi: [10.1105/tpc.114.132506](https://doi.org/10.1105/tpc.114.132506) PMID: [WOS:000350764700013](https://pubmed.ncbi.nlm.nih.gov/25000350764700013/).
8. Lasky JR, Des Marais DL, McKay JK, Richards JH, Juenger TE, Keitt TH. Characterizing genomic variation of *Arabidopsis thaliana*: the roles of geography and climate. *Mol Ecol*. 2012; 21(22):5512–29. doi: [10.1111/j.1365-294X.2012.05709.x](https://doi.org/10.1111/j.1365-294X.2012.05709.x) PMID: [WOS:000310583800013](https://pubmed.ncbi.nlm.nih.gov/25000310583800013/).
9. Clemens S. Developing tools for phytoremediation: towards a molecular understanding of plant metal tolerance and accumulation. *Int J Occup Med Environ Health*. 2001; 14(3):235–9. PMID: [11764851](https://pubmed.ncbi.nlm.nih.gov/11764851/).

10. Bomblies K, Weigel D. Arabidopsis and relatives as models for the study of genetic and genomic incompatibilities. *Philos T R Soc B*. 2010; 365(1547):1815–23. doi: [10.1098/rstb.2009.0304](https://doi.org/10.1098/rstb.2009.0304) PMID: [WOS:000277208600008](https://pubmed.ncbi.nlm.nih.gov/200277208600008/).
11. Page DR, Grossniklaus L. The art and design of genetic screens: *Arabidopsis thaliana*. *Nature Reviews Genetics*. 2002; 3(2):124–36. doi: [10.1038/nrg730](https://doi.org/10.1038/nrg730) PMID: [WOS:000173652700013](https://pubmed.ncbi.nlm.nih.gov/1173652700013/).
12. Snape JW, Lawrence MJ. Breeding System of *Arabidopsis-Thaliana*. *Heredity*. 1971; 27(Oct):299–302. doi: [10.1038/hdy.1971.91](https://doi.org/10.1038/hdy.1971.91) PMID: [WOS:A1971K911100011](https://pubmed.ncbi.nlm.nih.gov/11971100011/).
13. Kinoshita T, Ikeda Y, Ishikawa R. Genomic imprinting: A balance between antagonistic roles of parental chromosomes. *Seminars in Cell & Developmental Biology*. 2008; 19(6):574–9. doi: [10.1016/j.semcd.2008.07.018](https://doi.org/10.1016/j.semcd.2008.07.018) PMID: [WOS:000262670700014](https://pubmed.ncbi.nlm.nih.gov/11971100014/).
14. Ravi M, Marimuthu MPA, Tan EH, Maheshwari S, Henry IM, Marin-Rodriguez B, et al. A haploid genetics toolbox for *Arabidopsis thaliana*. *Nature Communications*. 2014; 5:5334. doi: [10.1038/ncomms6334](https://doi.org/10.1038/ncomms6334) PMID: [WOS:000344062400002](https://pubmed.ncbi.nlm.nih.gov/11971100002/).
15. Jiang CF, Mithani A, Belfield EJ, Mott R, Hurst LD, Harberd NP. Environmentally responsive genome-wide accumulation of de novo *Arabidopsis thaliana* mutations and epimutations. *Genome Research*. 2014; 24(11):1821–9. doi: [10.1101/gr.177659.114](https://doi.org/10.1101/gr.177659.114) PMID: [WOS:000344442000010](https://pubmed.ncbi.nlm.nih.gov/11971100010/).
16. *Arabidopsis Genome I*. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*. 2000; 408(6814):796–815. doi: [10.1038/35048692](https://doi.org/10.1038/35048692) PMID: [11130711](https://pubmed.ncbi.nlm.nih.gov/11130711/).
17. Roeder AH, Yanofsky MF. Fruit development in *Arabidopsis*. *Arabidopsis Book*. 2006; 4:e0075. doi: [10.1199/tab.0075](https://doi.org/10.1199/tab.0075) PMID: [22303227](https://pubmed.ncbi.nlm.nih.gov/22303227/); PubMed Central PMCID: [PMC3243326](https://pubmed.ncbi.nlm.nih.gov/PMC3243326/).
18. Eshed Y, Baum SF, Perea JV, Bowman JL. Establishment of polarity in lateral organs of plants. *Curr Biol*. 2001; 11(16):1251–60. PMID: [11525739](https://pubmed.ncbi.nlm.nih.gov/11525739/).
19. Bowman JL, Smyth DR. CRABS CLAW, a gene that regulates carpel and nectary development in *Arabidopsis*, encodes a novel protein with zinc finger and helix-loop-helix domains. *Development*. 1999; 126(11):2387–96. PMID: [10225998](https://pubmed.ncbi.nlm.nih.gov/10225998/).
20. Ferrandiz C. Regulation of fruit dehiscence in *Arabidopsis*. *J Exp Bot*. 2002; 53(377):2031–8. PMID: [12324527](https://pubmed.ncbi.nlm.nih.gov/12324527/).
21. Robles P, Pelaz S. Flower and fruit development in *Arabidopsis thaliana*. *Int J Dev Biol*. 2005; 49(5–6):633–43. doi: [10.1387/ijdb.052020pr](https://doi.org/10.1387/ijdb.052020pr) PMID: [16096970](https://pubmed.ncbi.nlm.nih.gov/16096970/).
22. Managbanag J, Witten TM, Bonchev D, Fox LA, Tsuchiya M, Kennedy BK, et al. Shortest-path network analysis is a useful approach toward identifying genetic determinants of longevity. *PLoS ONE*. 2008; 3(11):e3802. doi: [10.1371/journal.pone.0003802](https://doi.org/10.1371/journal.pone.0003802) PMID: [19030232](https://pubmed.ncbi.nlm.nih.gov/19030232/)
23. Chen L, Yang J, Huang T, Kong XY, Lu L, Cai YD. Mining for novel tumor suppressor genes using a shortest path approach. *Journal of Biomolecular Structure and Dynamics*. 2016; 34(3):664–75. doi: [10.1080/07391102.2015.1042915](https://doi.org/10.1080/07391102.2015.1042915) PMID: [26209080](https://pubmed.ncbi.nlm.nih.gov/26209080/)
24. Chen L, Xing Z, Huang T, Shu Y, Huang G, Li H-P. Application of the shortest path algorithm for the discovery of breast cancer related genes. *Current Bioinformatics*. 2016; 11(1):51–8.
25. Zhang J, Yang J, Huang T, Shu Y, Chen L. Identification of novel proliferative diabetic retinopathy related genes on protein-protein interaction network. *Neurocomputing*. 2016.
26. Chen L, Zhang YH, Huang T, Cai YD. Identifying novel protein phenotype annotations by hybridizing protein-protein interactions and protein sequence similarities. *Molecular Genetics and Genomics*. 2016; 291(2):913–34. doi: [10.1007/s00438-015-1157-9](https://doi.org/10.1007/s00438-015-1157-9) PMID: [26728152](https://pubmed.ncbi.nlm.nih.gov/26728152/)
27. Gui T, Dong X, Li R, Li Y, Wang Z. Identification of Hepatocellular Carcinoma-Related Genes with a Machine Learning and Network Analysis. *Journal of Computational Biology*. 2015; 22(1):63–71. doi: [10.1089/cmb.2014.0122](https://doi.org/10.1089/cmb.2014.0122) PMID: [25247452](https://pubmed.ncbi.nlm.nih.gov/25247452/)
28. Li B-Q, Huang T, Liu L, Cai Y-D, Chou K-C. Identification of Colorectal Cancer Related Genes with mRMR and Shortest Path in Protein-Protein Interaction Network. *PLoS One*. 2012; 7(4):e33393. doi: [10.1371/journal.pone.0033393](https://doi.org/10.1371/journal.pone.0033393) PMID: [22496748](https://pubmed.ncbi.nlm.nih.gov/22496748/)
29. Chen L, Chu C, Lu J, Kong XY, Huang T, Cai YD. A computational method for the identification of new candidate carcinogenic and non-carcinogenic chemicals. *Molecular BioSystems*. 2015; 11(9):2541–50. doi: [10.1039/c5mb00276a](https://doi.org/10.1039/c5mb00276a) PMID: [26194467](https://pubmed.ncbi.nlm.nih.gov/26194467/)
30. Avraham S, Tung CW, Ilic K, Jaiswal P, Kellogg EA, McCouch S, et al. The Plant Ontology Database: a community resource for plant structure and developmental stages controlled vocabulary and annotations. *Nucleic Acids Res*. 2008; 36(Database issue):D449–54. Epub 2008/01/17. doi: [10.1093/nar/gkm908](https://doi.org/10.1093/nar/gkm908) PMID: [18194960](https://pubmed.ncbi.nlm.nih.gov/18194960/); PubMed Central PMCID: [PMC3238838](https://pubmed.ncbi.nlm.nih.gov/PMC3238838/).
31. Hu L, Huang T, Shi X, Lu WC, Cai YD, Chou KC. Predicting functions of proteins in mouse based on weighted protein-protein interaction network and protein hybrid properties. *PLoS One*. 2011; 6(1):e14556. Epub 2011/02/02. doi: [10.1371/journal.pone.0014556](https://doi.org/10.1371/journal.pone.0014556) PMID: [21283518](https://pubmed.ncbi.nlm.nih.gov/21283518/); PubMed Central PMCID: [PMC3023709](https://pubmed.ncbi.nlm.nih.gov/PMC3023709/).

32. Sharan R, Ulitsky I, Shamir R. Network-based prediction of protein function. *Mol Syst Biol.* 2007; 3:88. Epub 2007/03/14. PMID: [17353930](#); PubMed Central PMCID: PMC1847944.
33. Ng KL, Ciou JS, Huang CH. Prediction of protein functions based on function-function correlation relations. *Comput Biol Med.* 2010; 40(3):300–5. doi: [10.1016/j.complbiomed.2010.01.001](#) PMID: [ISI:000276592400008](#).
34. Gao P, Wang QP, Chen L, Huang T. Prediction of Human Genes Regulatory Functions Based on Protein-protein Interaction Network. *Protein and Peptide Letters.* 2012; 19(9):910–6. PMID: [22486617](#)
35. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, et al. STRING v9. 1: protein-protein interaction networks, with increased coverage and integration. *Nucleic acids research.* 2013; 41(D1):D808–D15.
36. Moon H, Lee B, Choi G, Shin D, Prasad DT, Lee O, et al. NDP kinase 2 interacts with two oxidative stress-activated MAPKs to regulate cellular redox state and enhances multiple stress tolerance in transgenic plants. *Proc Natl Acad Sci U S A.* 2003; 100(1):358–63. doi: [10.1073/pnas.252641899](#) PMID: [12506203](#); PubMed Central PMCID: PMCPMC140977.
37. Zimmermann S, Baumann A, Jaekel K, Marbach I, Engelberg D, Frohnmeyer H. UV-responsive genes of arabidopsis revealed by similarity to the Gcn4-mediated UV response in yeast. *J Biol Chem.* 1999; 274(24):17017–24. PMID: [10358052](#).
38. Im YJ, Kim JI, Shen Y, Na Y, Han YJ, Kim SH, et al. Structural analysis of Arabidopsis thaliana nucleoside diphosphate kinase-2 for phytochrome-mediated light signaling. *J Mol Biol.* 2004; 343(3):659–70. doi: [10.1016/j.jmb.2004.08.054](#) PMID: [15465053](#).
39. Fujimori N, Suzuki N, Nakajima Y, Suzuki S. Plant DNA-damage repair/tolerance 100 protein repairs UV-B-induced DNA damage. *DNA repair.* 2014; 21:171–6. doi: [10.1016/j.dnarep.2014.05.009](#) PMID: [24951183](#).
40. Li CJ, Liu ZJ, Zhang QR, Wang RZ, Xiao LT, Ma H, et al. SKP1 is involved in abscisic acid signalling to regulate seed germination, stomatal opening and root growth in Arabidopsis thaliana. *Plant Cell Environ.* 2012; 35(5):952–65. doi: [10.1111/j.1365-3040.2011.02464.x](#) PMID: [WOS:000302541100008](#).
41. Anand A, Rojas CM, Tang YH, Mysore KS. Several components of SKP1/Cullin/F-box E3 ubiquitin ligase complex and associated factors play a role in Agrobacterium-mediated plant transformation. *New Phytologist.* 2012; 195(1):203–16. doi: [10.1111/j.1469-8137.2012.04133.x](#) PMID: [WOS:000304448500023](#).
42. Liu F, Ni W, Griffith ME, Huang Z, Chang C, Peng W, et al. The ASK1 and ASK2 genes are essential for Arabidopsis early development. *Plant Cell.* 2004; 16(1):5–20. doi: [10.1105/tpc.017772](#) PMID: [14688296](#); PubMed Central PMCID: PMCPMC301391.
43. Somers DE, Schultz TF, Milnamow M, Kay SA. ZEITLUPE encodes a novel clock-associated PAS protein from Arabidopsis. *Cell.* 2000; 101(3):319–29. PMID: [10847686](#).
44. Jarillo JA, Capel J, Tang RH, Yang HQ, Alonso JM, Ecker JR, et al. An Arabidopsis circadian clock component interacts with both CRY1 and phyB. *Nature.* 2001; 410(6827):487–90. doi: [10.1038/35068589](#) PMID: [WOS:000167583800046](#).
45. Malinova I, Kunz HH, Alseekh S, Herbst K, Fernie AR, Gierth M, et al. Reduction of the Cytosolic Phosphoglucomutase in Arabidopsis Reveals Impact on Plant Growth, Seed and Root Development, and Carbohydrate Partitioning. *Plos One.* 2014; 9(11):e112468. doi: [10.1371/journal.pone.0112468](#) PMID: [WOS:000345158700047](#).
46. Egli B, Kolling K, Kohler C, Zeeman SC, Streb S. Loss of Cytosolic Phosphoglucomutase Compromises Gametophyte Development in Arabidopsis. *Plant Physiol.* 2010; 154(4):1659–71. doi: [10.1104/pp.110.165027](#) PMID: [WOS:000284834000007](#).
47. Adwy W, Laxa M, Peterhansel C. A simple mechanism for the establishment of C2 -specific gene expression in Brassicaceae. *Plant J.* 2015; 84(6):1231–8. doi: [10.1111/tpj.13084](#) PMID: [26603271](#).
48. Hasse D, Andersson E, Carlsson G, Masloboy A, Hagemann M, Bauwe H, et al. Structure of the homodimeric glycine decarboxylase P-protein from Synechocystis sp. PCC 6803 suggests a mechanism for redox regulation. *J Biol Chem.* 2013; 288(49):35333–45. doi: [10.1074/jbc.M113.509976](#) PMID: [24121504](#); PubMed Central PMCID: PMCPMC3853282.
49. Chmielowska-Bak J, Lefevre I, Lutts S, Deckert J. Short term signaling responses in roots of young soybean seedlings exposed to cadmium stress. *J Plant Physiol.* 2013; 170(18):1585–94. doi: [10.1016/j.jplph.2013.06.019](#) PMID: [23942356](#).
50. Whitelam GC, Johnson E, Peng J, Carol P, Anderson ML, Cowl JS, et al. Phytochrome A null mutants of Arabidopsis display a wild-type phenotype in white light. *Plant Cell.* 1993; 5(7):757–68. doi: [10.1105/tpc.5.7.757](#) PMID: [8364355](#); PubMed Central PMCID: PMCPMC160314.

51. Oh S, Montgomery BL. Phytochrome-induced SIG2 expression contributes to photoregulation of phytochrome signalling and photomorphogenesis in *Arabidopsis thaliana*. *J Exp Bot*. 2013; 64(18):5457–72. doi: [10.1093/jxb/ert308](https://doi.org/10.1093/jxb/ert308) PMID: [24078666](https://pubmed.ncbi.nlm.nih.gov/24078666/); PubMed Central PMCID: PMC3871806.
52. Wang Q, Fristedt R, Yu XH, Chen ZG, Liu HT, Lee Y, et al. The gamma-Carbonic Anhydrase Subcomplex of Mitochondrial Complex I Is Essential for Development and Important for Photomorphogenesis of *Arabidopsis*. *Plant Physiol*. 2012; 160(3):1373–83. doi: [10.1104/pp.112.204339](https://doi.org/10.1104/pp.112.204339) PMID: [WOS:000310584200018](https://pubmed.ncbi.nlm.nih.gov/22000018/).
53. Sadanandom A, Adam E, Orosa B, Viczian A, Klose C, Zhang CJ, et al. SUMOylation of phytochrome-B negatively regulates light-induced signaling in *Arabidopsis thaliana* (vol 112, pg 11108, 2015). *Proceedings of the National Academy of Sciences of the United States of America*. 2015; 112(39):E5447–E8. doi: [10.1073/pnas.1517186112](https://doi.org/10.1073/pnas.1517186112) PMID: [WOS:000361941600013](https://pubmed.ncbi.nlm.nih.gov/26000013/).
54. Ronceret A, Gadea-Vacas J, Guillemot J, Lincker F, Delorme V, Lahmy S, et al. The first zygotic division in *Arabidopsis* requires de novo transcription of thymidylate kinase. *Plant J*. 2008; 53(5):776–89. doi: [10.1111/j.1365-3113X.2007.03372.x](https://doi.org/10.1111/j.1365-3113X.2007.03372.x) PMID: [WOS:000253484600007](https://pubmed.ncbi.nlm.nih.gov/18000007/).
55. Ramalingam J, Cruz CMV, Kukreja K, Chittoor JM, Wu JL, Lee SW, et al. Candidate Defense genes from rice, barley, and maize and their association with qualitative and quantitative resistance in rice. *Mol Plant Microbe In*. 2003; 16(1):14–24. doi: [10.1094/Mpmi.2003.16.1.14](https://doi.org/10.1094/Mpmi.2003.16.1.14) PMID: [WOS:000179860800002](https://pubmed.ncbi.nlm.nih.gov/12000002/).
56. Huang H, Quint M, Gray WM. The eta7/csn3-3 auxin response mutant of *Arabidopsis* defines a novel function for the CSN3 subunit of the COP9 signalosome. *PLoS One*. 2013; 8(6):e66578. doi: [10.1371/journal.pone.0066578](https://doi.org/10.1371/journal.pone.0066578) PMID: [23762492](https://pubmed.ncbi.nlm.nih.gov/23762492/); PubMed Central PMCID: PMC3676356.
57. Li S, Zachgo S. TCP3 interacts with R2R3-MYB proteins, promotes flavonoid biosynthesis and negatively regulates the auxin response in *Arabidopsis thaliana*. *Plant J*. 2013; 76(6):901–13. doi: [10.1111/tpj.12348](https://doi.org/10.1111/tpj.12348) PMID: [24118612](https://pubmed.ncbi.nlm.nih.gov/24118612/).
58. Tan M, Davis SW, Saunders TL, Zhu Y, Sun Y. RBX1/ROC1 disruption results in early embryonic lethality due to proliferation failure, partially rescued by simultaneous loss of p27. *Proceedings of the National Academy of Sciences of the United States of America*. 2009; 106(15):6203–8. doi: [10.1073/pnas.0812425106](https://doi.org/10.1073/pnas.0812425106) PMID: [WOS:000265174600033](https://pubmed.ncbi.nlm.nih.gov/19000033/).
59. Bernhardt A, Mooney S, Hellmann H. *Arabidopsis* DDB1a and DDB1b are critical for embryo development. *Planta*. 2010; 232(3):555–66. doi: [10.1007/s00425-010-1195-9](https://doi.org/10.1007/s00425-010-1195-9) PMID: [WOS:000280081800001](https://pubmed.ncbi.nlm.nih.gov/20000001/).
60. Seo KI, Lee JH, Nezames CD, Zhong SW, Song E, Byun MO, et al. ABD1 Is an *Arabidopsis* DCAF Substrate Receptor for CUL4-DDB1-Based E3 Ligases That Acts as a Negative Regulator of Abscisic Acid Signaling. *Plant Cell*. 2014; 26(2):695–711. doi: [10.1105/tpc.113.119974](https://doi.org/10.1105/tpc.113.119974) PMID: [WOS:000333688300015](https://pubmed.ncbi.nlm.nih.gov/24000015/).
61. Lee JH, Terzaghi W, Gusmaroli G, Charron JBF, Yoon HJ, Chen HD, et al. Characterization of *Arabidopsis* and rice DWD proteins and their roles as substrate receptors for CUL4-RING E3 ubiquitin ligases. *Plant Cell*. 2008; 20(1):152–67. doi: [10.1105/tpc.107.055418](https://doi.org/10.1105/tpc.107.055418) PMID: [WOS:000256414400013](https://pubmed.ncbi.nlm.nih.gov/18000013/).
62. Gao LY, Yang SZ, Zhu YY, Zhang JF, Zhuo M, Miao M, et al. The tomato DDI2, a PCNA ortholog, associating with DDB1-CUL4 complex is required for UV-damaged DNA repair and plant tolerance to UV stress. *Plant Sci*. 2015; 235:101–10. doi: [10.1016/j.plantsci.2015.03.007](https://doi.org/10.1016/j.plantsci.2015.03.007) PMID: [WOS:000354591700010](https://pubmed.ncbi.nlm.nih.gov/25000010/).
63. Chung K, Tasaka M. RPT2a, a 26S proteasome AAA-ATPase, is Directly Involved in *Arabidopsis* CC-NBS-LRR Protein uni-1D-Induced Signaling Pathways. *Plant Cell Physiol*. 2011; 52(9):1657–64. doi: [10.1093/pcp/pcr099](https://doi.org/10.1093/pcp/pcr099) PMID: [WOS:000294996200016](https://pubmed.ncbi.nlm.nih.gov/22000016/).
64. Kurepa J, Wang S, Li Y, Zaitlin D, Pierce AJ, Smalle JA. Loss of 26S Proteasome Function Leads to Increased Cell Size and Decreased Cell Number in *Arabidopsis* Shoot Organs. *Plant Physiol*. 2009; 150(1):178–89. doi: [10.1104/pp.109.135970](https://doi.org/10.1104/pp.109.135970) PMID: [WOS:000265661700016](https://pubmed.ncbi.nlm.nih.gov/19000016/).
65. Zarate X, Henderson DC, Phillips KC, Lake AD, Galbraith DW. Development of high-yield autofluorescent protein microarrays using hybrid cell-free expression with combined *Escherichia coli* S30 and wheat germ extracts. *Proteome science*. 2010; 8:32. doi: [10.1186/1477-5956-8-32](https://doi.org/10.1186/1477-5956-8-32) PMID: [WOS:000280282700001](https://pubmed.ncbi.nlm.nih.gov/20000001/).
66. Ye W, Hu S, Wu L, Ge C, Cui Y, Chen P, et al. (), encoding a nucleoside diphosphate kinase 2 (OsNDPK2), regulates chloroplast development and abiotic stress response in rice (*L.*). *Mol Breed*. 2016; 36:57. doi: [10.1007/s11032-016-0479-6](https://doi.org/10.1007/s11032-016-0479-6) PMID: [27212889](https://pubmed.ncbi.nlm.nih.gov/27212889/); PubMed Central PMCID: PMC4851688.
67. Seong ES, Guo J, Kim YH, Cho JH, Lim CK, Hyun Hur J, et al. Regulations of marker genes involved in biotic and abiotic stress by overexpression of the AtNDPK2 gene in rice. *Biochem Biophys Res Commun*. 2007; 363(1):126–32. doi: [10.1016/j.bbrc.2007.08.147](https://doi.org/10.1016/j.bbrc.2007.08.147) PMID: [17826739](https://pubmed.ncbi.nlm.nih.gov/17826739/).

68. Copetti D, Zhang J, El Baidouri M, Gao D, Wang J, Barghini E, et al. RITE database: a resource database for genus-wide rice genomics and evolutionary biology. *BMC Genomics*. 2015; 16:538. doi: [10.1186/s12864-015-1762-3](https://doi.org/10.1186/s12864-015-1762-3) PMID: [26194356](https://pubmed.ncbi.nlm.nih.gov/26194356/); PubMed Central PMCID: PMC4508813.
69. Kumar B, Chaudhuri DK. Isolation, partial characterisation of antithiamine factor present in rice-bran and its effect on TPP-transketolase system and *Staphylococcus aureus*. *Int J Vitam Nutr Res*. 1976; 46(2):154–9. PMID: [1032626](https://pubmed.ncbi.nlm.nih.gov/1032626/).
70. Li B, Ji L, Wu Y, Hao X. Isolation, prokaryotic expression and activity analysis of thymidylate kinase (tmk) gene from Phytoplasma of wheat blue dwarf. *Wei Sheng Wu Xue Bao*. 2008; 48(6):739–44. PMID: [18720837](https://pubmed.ncbi.nlm.nih.gov/18720837/).
71. Chen W, Li Y, Wang Q, Wang N, Wu YF. Comparative Genome Analysis of Wheat Blue Dwarf Phytoplasma, an Obligate Pathogen That Causes Wheat Blue Dwarf Disease in China. *Plos One*. 2014; 9(5): e96436. doi: [10.1371/journal.pone.0096436](https://doi.org/10.1371/journal.pone.0096436) PMID: [WOS:000336656000076](https://pubmed.ncbi.nlm.nih.gov/WOS:000336656000076/).
72. Goins GD, Yorio NC, Sanwo MM, Brown CS. Photomorphogenesis, photosynthesis, and seed yield of wheat plants grown under red light-emitting diodes (LEDs) with and without supplemental blue lighting. *Journal of Experimental Botany*. 1997; 48(312):1407–13. doi: [10.1093/jxb/48.7.1407](https://doi.org/10.1093/jxb/48.7.1407) PMID: [WOS:A1997XT43800008](https://pubmed.ncbi.nlm.nih.gov/WOS:A1997XT43800008/).
73. Zou Q, Li J, Song L, Zeng X, Wang G. Similarity computation strategies in the microRNA-disease network: a survey. *Briefings in Functional Genomics*. 2016; 15(1):55–64. PMID: [10.1093/bfpg/elv024](https://pubmed.ncbi.nlm.nih.gov/10.1093/bfpg/elv024/) doi: [10.1093/bfpg/elv024](https://doi.org/10.1093/bfpg/elv024)
74. Liu Y, Zeng X, He Z, Zou Q. Inferring microRNA-disease associations by random walk on a heterogeneous network with multiple data sources. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*. 2016. doi: [10.1109/TCBB.2016.2550432](https://doi.org/10.1109/TCBB.2016.2550432)
75. Shen Y, Han YJ, Kim JI, Song PS. Arabidopsis nucleoside diphosphate kinase-2 as a plant GTPase activating protein. *BMB reports*. 2008; 41(9):645–50. WOS:PMID: [000259782600005](https://pubmed.ncbi.nlm.nih.gov/000259782600005/).