# Identifying novel fruit-related genes in *Arabidopsis thaliana* based on the random walk with restart algorithm

**Yunhua Zhang**[1]*, **Li Dai**[1], **Ying Liu**[2], **YuHang Zhang**[3], **ShaoPeng Wang**[4]

**1** School of Resources and Environment, Anhui Agricultural University, Hefei, China, **2** School of Forestry and Landscape Architecture, Anhui Agricultural University, Hefei, China, **3** Institute of Health Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China, **4** School of Life Sciences, Shanghai University, Shanghai, China

* yunhua9681@163.com

## Abstract

Fruit is essential for plant reproduction and is responsible for protection and dispersal of seeds. The development and maturation of fruit is tightly regulated by numerous genetic factors that respond to environmental and internal stimulation. In this study, we attempted to identify novel fruit-related genes in a model organism, *Arabidopsis thaliana*, using a computational method. Based on validated fruit-related genes, the random walk with restart (RWR) algorithm was applied on a protein-protein interaction (PPI) network using these genes as seeds. The identified genes with high probabilities were filtered by the permutation test and linkage tests. In the permutation test, the genes that were selected due to the structure of the PPI network were discarded. In the linkage tests, the importance of each candidate gene was measured from two aspects: (1) its functional associations with validated genes and (2) its similarity with validated genes on gene ontology (GO) terms and KEGG pathways. Finally, 255 inferred genes were obtained, subsequent extensive analysis of important genes revealed that they mainly contribute to ubiquitination (UBQ9, UBQ8, UBQ11, UBQ10), serine hydroxymethyl transfer (SHM7, SHM5, SHM6) or glycol-metabolism (HXKL2_ARATH, CSY5, GAPCP1), suggesting essential roles during the development and maturation of fruit in *Arabidopsis thaliana*.

## Introduction

Fruit, as specialized seed-bearing structures, are designed to protect the reproductive organs of plants [1]. During the development and maturation of seeds, fruit has been confirmed to play a specific role. Based on recent publications, there are two main functions of fruit that may contribute to the reproductive processes of plants [2]. First, fruits protect seeds. For most angiosperms, fruits serve as a solid physical barrier between seeds and the external environment, providing a reliable shelter for seed development and maturation [2, 3]. Second, some fruits may contribute to the dispersal of mature seeds. Not all mature seeds of a plant can grow into fertile plants. During the germination and growing processes of a mature seed, the

surrounding environment is highly significant [4, 5]. During transfer of the seed to a proper environment for further germination and growth, fruits play a crucial role [5]. Take coffee as an example. The solid fruit of coffee berries protect the seed inside from the diverse external environment [6, 7]. Digestion by a specific civet cat (*Paradoxurus hermaphroditus*) can destroy the fruit and expose the seeds, which also may complete the dispersal process [8]. Considering fruit is crucial for the plant reproduction, it is important to identify fruit-associated regulatory factors. Among such factors, specific intrinsic factors, especially genetic factors, have been confirmed to play an irreplaceable role during the development and maturation of fruit.

*Arabidopsis thaliana* is an annual plant native to Eurasia, and has a height of 20–25 cm [9]. First described in 1577 by Johannes Thal, *Arabidopsis thaliana* has unique advantages for use as a model plant. At first, *Arabidopsis thaliana* has an appropriate size (7–40 cm), which not only is suitable for morphologic observation but also enables large-scale indoor cultivation [10]. Furthermore, the reproductive capacity of Arabidopsis is robust, enabling scientists to obtain many plant seedlings within a short time [10]. Further, as a self-pollinated plant, the fertilization process of *Arabidopsis thaliana* can be easily interfered with and controlled to meet the needs of biological experiments, avoiding other external environmental factors and making it a perfect model organism for genetics research [11]. Finally, the whole genome of *Arabidopsis thaliana*, the smallest among cruciferous plants, only contains five chromosomes and 100 million base pairs that have been completely sequenced [12]. Considering these advantages, *Arabidopsis thaliana* is an optimal model organism and has contributed greater understanding in genetics and botany.

As we have analyzed above, fruit associated biological processes are regulated by multiple internal and external regulatory factors in multiple plant subtypes, including *Arabidopsis thaliana*, among which genetic factors play a specific critical role. Considering the advantages of *Arabidopsis thaliana* as a typical model plant, as we have analyzed above, it is reasonable to take such plant as a model for further study on fruit associated genes. In *Arabidopsis thaliana*, a self-pollinated plant, various genes contribute to regulation of fruit development. For example, *KANADI1* and *KANADI2* participate in the development of lateral organs [13]. However, it is expensive and time-consuming to identify functional fruit-associated genes in *Arabidopsis thaliana* by traditional experimental methods. On the other hand, with the development of computer technology and its successful application in the fields of biology, medicine [14–19], it becomes possible to develop reliable computational methods for the identification of fruit-associated genes.

Here, based on the validated fruit-related genes in *Arabidopsis thaliana*, we tried to identify novel fruit-related genes using computational techniques. Up to now, several network methods have been developed to identify novel disease genes [15, 20–25]. Some of these methods employed the classic network algorithms, such as random walk with restart (RWR) algorithm [26, 27] and shortest path algorithm [28]. These algorithms gave novel directions for identification of functional fruit-associated genes. Recently, Zhu *et al*. built a method that used the shortest path algorithm as the basic search algorithm to identify novel fruit-associated genes in *Arabidopsis thaliana* [29]. In this study, we adopted the RWR algorithm to construct the method. In detail, it was executed on a protein-protein interaction (PPI) network, which was constructed using PPI information in STRING [30], with validated genes as seed nodes. Genes with high probabilities were selected. To identify false positives, the permutation test and linkage tests were built to screen out essential genes. The permutation test can discard genes selected due to the structure of the PPI network. In the linkage tests, we measured the importance of each candidate gene from two points: (1) its functional associations with validated genes and (2) its similarity with validated genes on gene ontology (GO) terms [31] and Kyoto encyclopedia of genes and genomes (KEGG) pathways [32]. A group of inferred genes were

accessed, some of which were extensively analyzed. Our results indicate that these genes may participate in fruit-associated biological processes of *Arabidopsis thaliana*.

## Materials and methods

### Dataset

To obtain the validated fruit-related genes in *Arabidopsis thaliana*, we first accessed fruit-related PO terms. A file named plant_ontology.obo (accessed on March 24, 2005) was downloaded from Plant Ontology (PO, http://www.plantontology.org/download) [33]. This file provided easy-to-retrieve structures of PO terms. Terms containing fruit (PO: 0009001) and its children terms (PO: 0004707, PO: 0008001, PO: 0004536, PO: 0004535, PO: 0008002, PO: 0025268, PO: 0000033, PO: 0008003, PO: 0009087, and PO: 0009084) were extracted and regarded as fruit-related PO terms. The descriptions of these PO terms are listed in Table 1. Accordingly, 994 genes annotated by these PO terms were accessed and considered as the validated fruit-related genes. The IDs of these 994 genes are provided in S1 Table.

### PPI network

Because the functions of proteins need some factors to regulate, intercellular and intracellular proteins rarely execute their functions alone. Thus, the PPIs were essential for the normal metabolism of organisms. Some computational methods have been developed to identify novel PPIs [34–37], which can yield more abundant PPIs to investigate related problems. The accumulated information of PPIs can be used to construct a large PPI network, which is helpful to investigate different properties of proteins, such as protein functions [38–40], relationship with different diseases [21, 22, 41–45].

In this study, we used the PPIs of *Arabidopsis thaliana* reported in STRING [30] (http://string-db.org/, Version 9.1), a well-known public database collecting PPIs of several organisms, to construct the PPI network. PPIs in STRING are derived from genomic context, high throughput experiments, (conserved) co-expression, and previous knowledge, indicating they can measure both the direct (physical) and indirect (functional) associations between proteins. To retrieve the PPIs of *Arabidopsis thaliana*, a file 'protein.links.v9.1.txt.gz' was downloaded from STRING, which contains PPIs of 1,133 organisms covering 5,214,234 proteins. Because '3702' is the organism code of *Arabidopsis thaliana* in STRING, lines starting with '3702' in the obtained file were extracted, obtaining 3,123,482 PPIs covering 25,123 proteins of *Arabidopsis thaliana*. In each PPI, there are two Ensembl IDs and a score ranging between 150 and 999.

**Table 1. The information of fruit-related PO terms.**

| PO term ID | Description |
|---|---|
| PO:0009001 | fruit |
| PO:0004707 | fruit dehiscence zone |
| PO:0008001 | fruit distal end |
| PO:0004536 | fruit pedicel |
| PO:0004535 | fruit placenta |
| PO:0008002 | fruit proximal end |
| PO:0025268 | fruit septum |
| PO:0000033 | fruit valve |
| PO:0008003 | fruit vascular system |
| PO:0009087 | mesocarp |
| PO:0009084 | pericarp |

https://doi.org/10.1371/journal.pone.0177017.t001

For formulation, the score of a PPI with proteins $p_a$ and $p_b$ was denoted by $S_I(p_a, p_b)$. The constructed PPI network $G$ defined 25,123 proteins as nodes and each edge represented one PPI, i.e., two nodes were adjacent if and only if their corresponding proteins can constitute a PPI. In addition, the score of each interaction was assigned to the corresponding edge as its weight.

## RWR algorithm

The RWR algorithm [46, 47] is a classic ranking algorithm. This algorithm stimulates a walker that starts from a seed node or some seed nodes and randomly moves in a network. Here, 994 fruit-related genes were deemed as seed nodes, and the RWR algorithm was applied on the PPI network $G$ to discover possible novel fruit-related genes.

Before executing the RWR algorithm on $G$, an initialization vector $P_0$ was constructed containing 25,123 components that corresponded to 25,123 nodes in the PPI network $G$. In $P_0$, the component corresponding to the validated fruit-related genes was set to 1/994 and others were set to zero. The RWR algorithm repeatedly updates this vector and denotes $P_i$ as the vector after the $i$-th round has been done. The updating rule is as follows:

$$P_{i+1} = A^T P_i + cP_0 \qquad (1)$$

where $A$ is the column normalized adjacency matrix (the sum of members in each column equals to one) of the PPI network $G$ and $c$ is the restart probability (it was set to 0.8 in this study). The updating procedure stops until the vector $P_i$ becomes stable, which is measured by the condition of $\| P_{i+1} - P_i \|_{L1} < 10^{-6}$. The vector $P_{i+1}$ is outputted, which indicates the probabilities of all nodes (genes) to be fruit-related genes.

Clearly, genes receiving larger probabilities are more likely to be the potential fruit-related genes. To avoid omitting too many potential genes, we set a threshold of $10^{-5}$, which was used in another study [24] to select possible genes, i.e., genes with probabilities larger than $10^{-5}$ were selected for further analysis and were called RWR genes for convenience.

## Permutation test

In Section "RWR algorithm", the RWR algorithm was applied on the PPI network $G$ using validated fruit-related genes as seed nodes, producing some RWR genes. However, not all these genes are tightly associated with the fruits of *Arabidopsis thaliana*. Some RWR genes were selected because of the structure of the PPI network, i.e., the structure of the network can influence the utility of the RWR algorithm. To discard these genes, the permutation test [43, 48–50] was designed as follows.

1. 1,000 node sets, formulated as $S_1, S_2, S_3 \ldots, S_{1000}$, were randomly constructed, and each set contained 994 nodes in the PPI network $G$;

2. For each set, the RWR algorithm was applied on the PPI network $G$ using nodes in the set as seed nodes;

3. For each RWR gene, there were 1,000 probabilities produced by $S_1, S_2, S_3, \ldots, S_{1000}$ and one probability yielded by validated fruit-related genes. Accordingly, a P-value was calculated for each RWR gene $g$, which was defined as

$$P - value(g) = \Theta/1000 \qquad (2)$$

where $\Theta$ is the number of node sets on which the probabilities of $g$ were larger than that on validated fruit-related genes. Clearly, a RWR gene with a high P-value indicates that it is not specific for fruit because several randomly produced sets can discover it. Thus, we

should select RWR genes with low P-values. Because 0.05 is always selected as a cutoff of the significance level of statistical tests, it was set to be the threshold of the P-value. Thus, RWR genes with P-values less than 0.05 were selected, and the selected genes were called candidate genes for convenience.

## Linkage tests

As described in Section "Permutation test", some candidate genes with P-values less than 0.05 were selected. These genes are deemed to have more or less associations with the fruit of *Arabidopsis thaliana*. This section built two linkage tests that can identify candidate genes with close relation to the fruit of *Arabidopsis thaliana*.

It is known that proteins that can form a PPI are often share similar functions or located in same signal pathways [39, 40, 51, 52]. Accordingly, candidate genes that can interact with at least one validated fruit-related gene are more likely to be novel fruit-related genes. Furthermore, each interaction was assigned a score ranging between 150 and 999 to indicate its strength. Thus, we can further consider this score to measure the functional associations between candidate genes and validated fruit-related genes. For each candidate gene $g$, a measurement, called maximum interaction score (*MIS*), was calculated, which is defined as:

$$MIS(g) = \max\{S_I(g, g') : g' \text{ is a validated fruit} - \text{related gene}\} \qquad (3)$$

Clearly, a larger *MIS* for a candidate gene indicate tight interaction with at least one validated gene, and thus has a high likelihood to be a novel fruit-related gene. In the STRING database, 900 is the threshold of the highest confidence level of PPIs. Thus, this value was adopted as the threshold of *MIS* to screen out important candidate genes.

The GO terms [31] and KEGG pathways [32] are always utilized to elucidate and describe molecular functions, cellular components, biological and signal processes of genes. Each gene can be annotated by some GO terms or KEGG pathways. If a candidate gene exhibits similar GO terms or KEGG pathways with some validated fruit-related genes, it is more likely to be a novel fruit-related gene. To measure the similarity of candidate genes and validated genes on GO terms and KEGG pathways, the enrichment theory of GO terms and KEGG pathways was employed, based on which the relationship between a gene and GO terms or KEGG pathways can be formulated as a numeric vector [53–57]. For formulation, the vector yielded by the enrichment theory for a gene $g$ was denoted by $V(g)$. The similarity of two genes $g$ and $g'$ on GO terms and KEGG pathways can be measured by the proximity of their corresponding vectors, which was computed by

$$\Gamma(g, g') = \frac{V(g) \cdot V(g')}{\|V(g)\| \cdot \|V(g')\|} \qquad (4)$$

Obviously, a high outcome of Eq 4 indicates a close relationship between $g$ and $g'$. Like Eq 3, we can calculate a measurement, called maximum function score (*MFS*), for each candidate gene $g$, which was defined as

$$MFS(g) = \max\{\Gamma(g, g') : g' \text{ is a validated fruit} - \text{related gene}\} \qquad (5)$$

A large *MFS* for a given candidate gene implies significant overlap of GO terms or KEGG pathways with at least one validated fruit-related gene. 0.9 was set to be the threshold of *MFS* in this study, i.e., candidate genes with *MFS*s larger than 0.9 were selected.

Eventually, by considering these two linkage tests, candidate genes with *MIS*s no less than 900 and *MFS*s larger than 0.9 were selected as the putative genes in this study.
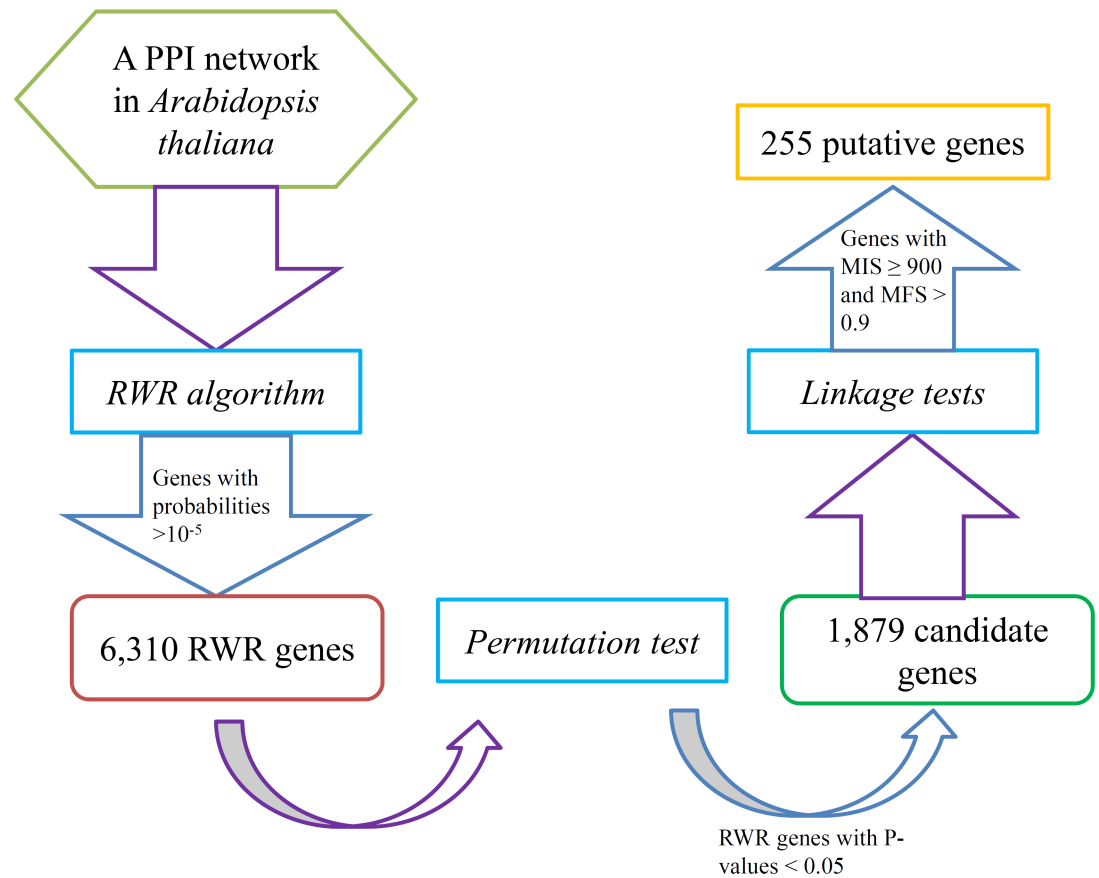
**Fig 1. The flowchart of the method for identifying novel fruit-related genes.**

https://doi.org/10.1371/journal.pone.0177017.g001

## Results

A flowchart ([Fig 1](#)) illustrates the entire method used for identifying novel fruit-related genes. This chart also shows the results yielded by each procedure.

Our method first applied the RWR algorithm to the PPI network constructed in Section "PPI network" using validated fruit-related genes as seed nodes. Each gene in the network received a probability that indicated the possibility of it being a novel fruit-related gene. To reduce the searching scope, we selected genes with probabilities larger than $10^{-5}$, obtaining 6,310 RWR genes. These genes and their probabilities yielded by the RWR algorithm are listed in S2 Table.

Among the 6,310 RWR genes, not all are related to the fruit of *Arabidopsis thaliana*. As mentioned in Section "Permutation test", some were selected due to the structure of the PPI network but had no relationship with the fruit of *Arabidopsis thaliana*. Thus, a permutation test was adopted to filter these types of RWR genes. A P-value was assigned to each RWR gene, which is also provided in S2 Table. Because 0.05 was selected as the criterion for statistical significance, we selected RWR genes with P-values less than 0.05, resulting in 1,879 candidate genes. These genes are available in S2 Table.

To further select essential fruit-related genes among the 1,879 candidate genes, two linkage tests were executed. Each candidate gene was assigned an MIS and an MFS, which are also available in S2 Table. The strict thresholds 900 and 0.9 for *MIS* and *MFS* were applied. 255

genes remained, which are listed in S3 Table. These genes were deemed to be significant to fruit of *Arabidopsis thaliana*. For convenience, they were called putative genes in this study.

## Discussion

Based on our computational method, we identified a group of putative genes (255 genes) that may directly or indirectly contribute to the development and maturation of the fruit in *Arabidopsis thaliana*. In another study, Zhu *et al.* adopted the shortest path algorithm to search possible fruit-related genes in *Arabidopsis thaliana* [29]. In fact, the shortest path algorithm always identifies possible genes using a pair of validated genes and collects all identified possible genes together as the candidate genes. Its principle is quite different from the RWR algorithm because RWR algorithm tries to search possible genes using the validated genes as a whole and diffusing the probabilities on validated genes to other possible genes. It is anticipated that the candidate genes obtained by these two algorithms are quite different. To prove this claim, we downloaded the identified genes in Zhu *et al.*'s study, totally 517 candidate genes. Of the 255 putative genes obtained in this study, 44 genes were also identified in Zhu *et al.*'s study and 211 genes were exclusively reported in our study (see S3 Table for detailed information). Less than one-fifth putative genes were identified in Zhu *et al.*'s study, which further proves the different influence of the shortest path algorithm and RWR algorithm for identifying possible fruit-related genes in *Arabidopsis thaliana*. Thus, the putative genes reported in this study can be an important supplement for the complete identification of fruit-related genes in *Arabidopsis thaliana*.

Among the 255 putative genes, 211 genes were identified in our study and not predicted in Zhu *et al.*' study. Of these 211 genes, some of these can be validated based upon results reported in recent publications, reflecting the accuracy and efficacy of our method. At this point in our investigation, we selected ten important putative genes (see Table 2), to analyze their relationship with the development and maturation of *Arabidopsis* fruit. The linkages between these ten putative genes and validated genes are illustrated in Fig 2. Intuitively, they all have strong associations with validated fruit-related genes, thereby inducing their close relationships with the development and maturation of the fruit in *Arabidopsis thaliana*.

Among the putative genes, various ubiquitin associated genes appear to contribute to fruit-associated biological processes. **UBQ9**, which is also known as AT5G37640 in *Arabidopsis thaliana*, has been mainly reported to contribute to the response to certain cytokines and the ubiquitin-dependent protein catabolic processes [58, 59]. Ubiquitin exists either attached to another protein or unanchored. Depending on the specific Lys site ubiquitin linked, anchored proteins have many specific functions, including lysosomal degradation, endocytosis and

**Table 2. The detailed information of ten important putative genes.**

| Ensembl ID | Gene Name | Protein name | Probability | P-value | MIS | MFS |
|---|---|---|---|---|---|---|
| AT5G37640 | UBQ9 | Polyubiquitin 9 | 7.47E-05 | <0.001 | 946 | 0.999 |
| AT3G09790 | UBQ8 | Polyubiquitin 8 | 7.49E-05 | <0.001 | 946 | 0.999 |
| AT4G05050 | UBQ11 | Polyubiquitin 11 | 8.00E-05 | <0.001 | 946 | 0.999 |
| AT4G05320 | UBQ10 | Polyubiquitin 10 | 8.37E-05 | <0.001 | 940 | 0.999 |
| AT1G36370 | SHM7 | Serine hydroxymethyltransferase 7 | 2.82E-05 | 0.001 | 994 | 0.997 |
| AT4G13890 | SHM5 | Serine hydroxymethyltransferase 5 | 3.11E-05 | 0.001 | 995 | 0.997 |
| AT1G22020 | SHM6 | Serine hydroxymethyltransferase 6 | 3.10E-05 | <0.001 | 994 | 0.997 |
| AT4G37840 | HXKL2_ARATH | Probable hexokinase-like 2 protein | 3.08E-05 | 0.007 | 966 | 0.997 |
| AT3G60100 | CSY5 | Citrate synthase 5 | 2.94E-05 | <0.001 | 980 | 0.997 |
| AT1G79530 | GAPCP1 | Glyceraldehyde-3-phosphate dehydrogenase | 2.98E-05 | 0.001 | 998 | 0.997 |

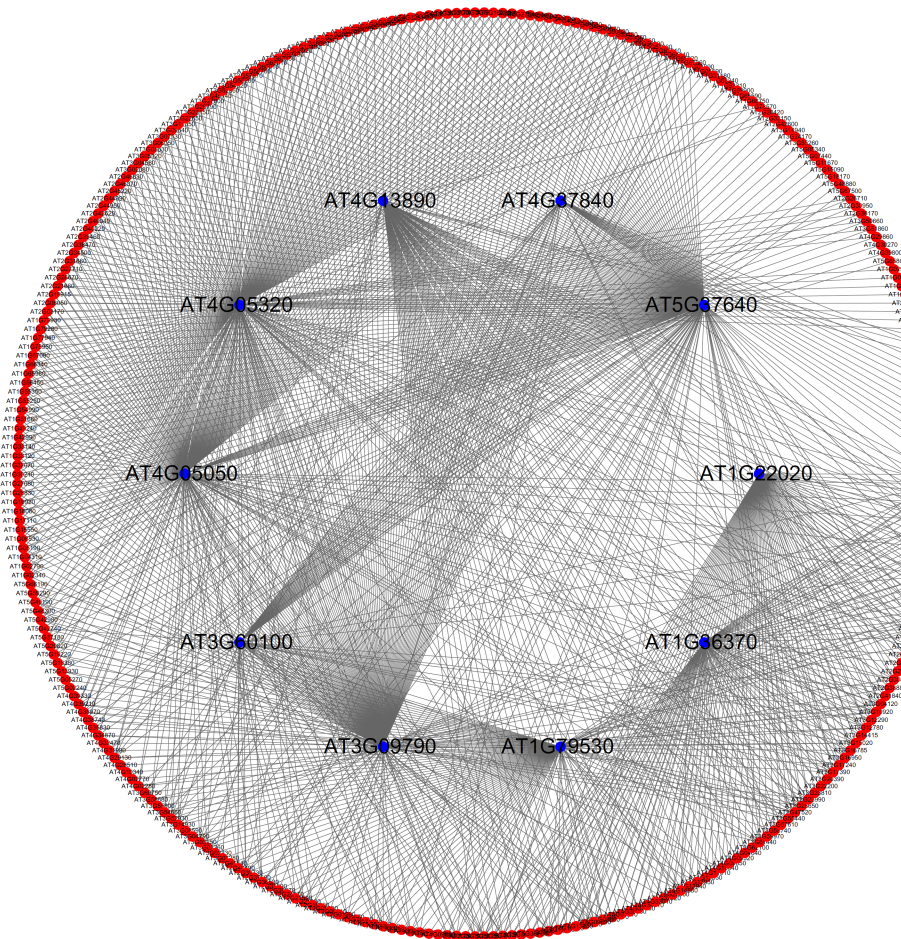https://doi.org/10.1371/journal.pone.0177017.t002

**Fig 2. The linkages between ten important putative genes and validated fruit-related genes that were extracted from the PPI network.** Red nodes represent validated fruit-related genes. Blue nodes represent putative genes.

https://doi.org/10.1371/journal.pone.0177017.g002

DNA damage response [60–62]. Ubiquitin-dependent proteins have been widely reported to contribute to the development and maturation of fruits in various plants, such as tomatoes, bananas and our model plant, *Arabidopsis thaliana* [63–65]. Recent publications confirmed that UBQ9 interacts with various core components of ubiquitin-dependent protein catabolic processes [58]. Therefore, the putative gene UBQ9 may contribute to the specific fruit-associated biological processes. Another putative gene, **UBQ8**, is known as AT3G09790 in *Arabidopsis thaliana*. It is the homologue of UBQ9. As mentioned above, ubiquitin-dependent proteins may mediate the degradation of specific target proteins, such as HOS1, MdCOP1, and ETO1, which may further promote the development and maturation of fruit in *Arabidopsis thaliana* [63–66]. Therefore, such genes may, such as its homologue UBQ9, contribute to specific fruit-associated metabolic processes. **UBQ11**, also known as AT4G05050, has also been confirmed as one of the poly-ubiquitin families which contribute to ERAD (endoplasmic reticulum-associated degradation) [67]. Recent publications confirm that, at least in rice, ERAD-associated biological processes may be directly related to the development and maturation of the fruit, implying its specific role in plant fruiting [68]. Considering that recent publications also identified ERAD-associated biological processes as a crucial regulator for reproduction (including fruits) of *Arabidopsis thaliana*, it is reasonable to regard the putative gene UBQ11

as a candidate fruit-associated gene [69]. The putative gene **UBQ10** encodes another specific ubiquitin-associated protein, also known as AT4G05320, involving ubiquitin-dependent protein catabolism, which we have analyzed above and confirmed its association with fruiting [70].

Apart from the ubiquitin genes analyzed above, we also obtained a group of functional enzymes. Three of them encode subtypes of serine hydroxy methyltransferases. **SHM7**, also known as AT1G36370, encodes the specific serine hydroxy methyltransferase 7, which further contributes to catalyzing the interconversion of serine and glycine [71]. Recent publications confirm that the main function of SHM7, pyridoxal phosphate binding, may contribute to the development and maturation of fruit in citrus, implying its potential role in various plant subtypes [72, 73]. Furthermore, a specific mutant screening of *Arabidopsis thaliana*, based on high-throughput HPLC-MS/MS assay has confirmed that, together with another enzyme threonine aldolase, the serine hydroxy methyltransferase may play a specific role during the development and maturation of fruit, affecting seed nutritional quality [74]. **SHM5** (AT4G13890), another crucial serine hydroxy methyltransferase, has also been predicted to be a candidate fruit-related gene in *Arabidopsis thaliana* [74]. Similar to SHM7 analyzed above, SHM5 has also been confirmed to interact with specific endogenous compounds in plants, such as 5-Methyltetrahydrofolate and 5-formyltetrahydrofolate, in one-carbon metabolism. Thus, SHM5 may further participate in the maturation of fruits in *Arabidopsis thaliana*, validating the specific relationship between this gene and fruit development [75–77]. **SHM6** (AT1G22020) also encodes a specific subtype of serine hydroxy methyltransferases. Although no publications confirmed that such genes contribute to the development or maturation of fruits, considering the crucial regulatory role of serine hydroxy methyltransferases in *Arabidopsis thaliana*, and validated interactions in this study, it is reasonable to regard SHM6 as a candidate functional fruit-related gene [78, 79].

Apart from such serine hydroxy methyltransferases, we also identified specific hexokinase-associated genes **HXKL2_ARATH** (probable hexokinase-like 2 protein) and HXL3 (hexokinase like 3), also known as AT4G37840. Recent publications have confirmed the complex functions of such genes in plants. Expression in citrus guard cells regulates specific sugar-sensing functions during fruit development and maturation, at least in citrus [80]. An earlier exploration confirmed that such genes may contribute to AtRGS1-medated sugar signaling in *Arabidopsis thaliana* [81]. Considering that sugar signaling contributes to the nutrient accumulation and freezing tolerance of fruit, and may affect the longevity of seeds, it is reasonable to conclude that as a functional sugar metabolism regulator, AT4G37840 may contribute to fruit- associated biological processes in *Arabidopsis thaliana* [2, 82, 83]. Another gene, **CSY5**, also known as AT3G60100, contribute to fruit-associated biological processes. Recent publications not only revealed the potential functions of citrate synthase during seed germination but also suggests a crucial role of in fruit development, especially at low temperature [84, 85]. Expressed in fruit and seedlings of *Arabidopsis thaliana*, the putative gene CSY5 may contribute to fruit associated biological processes [84]. **GAPCP1** (glyceraldehyde-3-phosphate dehydrogenase), also known as AT1G79530, is a candidate fruit-related gene in *Arabidopsis thaliana*. Involved in the plastid glycolytic pathway, this gene may contribute to the production of glycolytic energy in non-photosynthetic tissues, especially in fruits [86, 87]. Considering that the glycolytic energy in fruits is important in cellular metabolism and seed oil accumulation in the fruit of *Arabidopsis thaliana*, the putative gene GAPCP1 may be a crucial fruit-related gene.

As we have analyzed above, ten putative genes have been validated to participate in fruit associated biological processes of *Arabidopsis thaliana*. Based on recent publications and our analysis, we identified two crucial biological processes: ubiquitin- associated biological

processes and serine hydroxy methyltransferase-associated biological processes that may contribute to the development and maturation of fruit in *Arabidopsis thaliana*. The typical enrichment of some putative genes in such biological processes implies the role and underlying mechanisms of them for fruit development and maturation. The remaining putative genes are not discussed here but, may also be related to fruit-associated biological processes of *Arabidopsis thaliana*. We hope that other investigators will test this hypothesis.

## Conclusions

In this study, we utilized powerful computational techniques for identifying novel fruit-related genes in *Arabidopsis thaliana*, yielding 255 inferred genes. These genes may provide new directions to investigate the biological processes associated with fruiting in *Arabidopsis thaliana*. Furthermore, we believe that this method can be further applied to the recognition of various functional genes/proteins of multiple species, such as DNA-binding protein prediction [88], promoting the development of large-scale gene/protein function prediction and identification.

## Supporting information

**S1 Table. 994 validated fruit-related genes in *Arabidopsis thaliana* and their fruit-related PO terms.**
(DOCX)

**S2 Table. The 6,310 RWR genes with probabilities larger than $10^{-5}$.**
(DOCX)

**S3 Table. The 255 putative fruit-related genes.**
(DOCX)

## Acknowledgments

## Author Contributions

**Conceptualization:** YZ.

**Data curation:** YZ LD YL.

**Formal analysis:** YZ YHZ SPW.

**Funding acquisition:** YZ.

**Methodology:** YZ LD.

**Writing – original draft:** YZ LD YL.

**Writing – review & editing:** YHZ.

## References

1. Roeder AH, Yanofsky MF. Fruit development in Arabidopsis. Arabidopsis Book. 2006; 4:e0075. https://doi.org/10.1199/tab.0075 PMID: 22303227

2. Zhang GZ, Jin SH, Jiang XY, Dong RR, Li P, Li YJ, et al. Ectopic expression of UGT75D1, a glycosyltransferase preferring indole-3-butyric acid, modulates cotyledon development and stress tolerance in seed germination of Arabidopsis thaliana. Plant Mol Biol. 2016; 90(1–2):77–93. https://doi.org/10.1007/s11103-015-0395-x PMID: 26496910

3. Balanza V, Roig-Villanova I, Di Marzo M, Masiero S, Colombo L. Seed abscission and fruit dehiscence required for seed dispersal rely on similar genetic networks. Development. 2016; 143(18):3372–81. https://doi.org/10.1242/dev.135202 PMID: 27510967

4. Wang WQ, Song BY, Deng ZJ, Wang Y, Liu SJ, Moller IM, et al. Proteomic Analysis of Lettuce Seed Germination and Thermoinhibition by Sampling of Individual Seeds at Germination and Removal of Storage Proteins by Polyethylene Glycol Fractionation. Plant Physiol. 2015; 167(4):1332–U380. https://doi.org/10.1104/pp.15.00045 PMID: 25736209

5. Leeggangers HACF, Folta A, Muras A, Nap JP, Mlynarova L. Reduced seed germination in Arabidopsis over-expressing SWI/SNF2 ATPase genes. Physiol Plantarum. 2015; 153(2):318–26.

6. Pinto LVA, Da Silva EAA, Davide AC, De Jesus VAM, Toorop PE, Hilhorst HWM. Mechanism and control of Solanum lycocarpum seed germination. Ann Bot-London. 2007; 100(6):1175–87.

7. Agrawal P, Verma D, Daniell H. Expression of Trichoderma reesei beta-Mannanase in Tobacco Chloroplasts and Its Utilization in Lignocellulosic Woody Biomass Hydrolysis. Plos One. 2011; 6(12):e29302. e29302. https://doi.org/10.1371/journal.pone.0029302 PMID: 22216240

8. Jumhawan U, Putri SP, Yusianto, Marwani E, Bamba T, Fukusaki E. Selection of Discriminant Markers for Authentication of Asian Palm Civet Coffee (Kopi Luwak): A Metabolomics Approach. Journal of agricultural and food chemistry. 2013; 61(33):7994–8001. https://doi.org/10.1021/jf401819s PMID: 23889358

9. 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. https://doi.org/10.1111/j.1365-294X.2012.05709.x PMID: 22857709

10. Snape JW, Lawrence MJ. Breeding System of Arabidopsis-Thaliana. Heredity. 1971; 27(Oct):299–302.

11. 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.

12. 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. https://doi.org/10.1038/ncomms6334 PMID: 25358957

13. 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

14. Zeng X, Lin W, Guo M, Zou Q. A comprehensive overview and evaluation of circular RNA detection tools. PLoS Computational Biology. 2017.

15. Barabasi AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nat Rev Genet. 2011; 12(1):56–68. https://doi.org/10.1038/nrg2918 PMID: 21164525

16. Jalali S, Kapoor S, Sivadas A, Bhartiya D, Scaria V. Computational approaches towards understanding human long non-coding RNA biology. Bioinformatics. 2015; 31(14):2241–51. https://doi.org/10.1093/bioinformatics/btv148 PMID: 25777523

17. Su R, Zhang C, Pham TD, Davey R, Bischof L, Vallotton P, et al. Detection of tubule boundaries based on circular shortest path and polar-transformation of arbitrary shapes. Journal of microscopy. 2016; 264 (2):127–42. https://doi.org/10.1111/jmi.12421 PMID: 27172164

18. Wei L, Xing P, Shi G, Ji ZL, Zou Q. Fast prediction of protein methylation sites using a sequence-based feature selection technique. IEEE/ACM Trans Comput Biol Bioinform. 2017.

19. Wei L, Xing P, Tang J, Zou Q. PhosPred-RF: a novel sequence-based predictor for phosphorylation sites using sequential information only. IEEE Trans Nanobioscience. 2017.

20. Oliver S. Guilt-by-association goes global. Nature. 2000; 403(6770):601–3. PMID: 10688178

21. Chen L, Hao 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.

22. 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; 217:63–72.

23. Chen L, Yang J, Xing Z, Yuan F, Shu Y, Zhang Y, et al. An integrated method for the identification of novel genes related to oral cancer. PLoS ONE. 2017.

24. Guo W, Shang DM, Cao JH, Feng KY, He YC, Jiang Y, et al. Identifying and Analyzing Novel Epilepsy-Related Genes Using Random Walk with Restart Algorithm. Biomed Research International. 2017; 2017:6132436. https://doi.org/10.1155/2017/6132436 PMID: 28255556

25. Zeng X, Liao Y, Liu Y, Zou Q. Prediction and validation of disease genes using HeteSim Scores. IEEE/ACM Trans Comput Biol Bioinform. 2016.

26. Ham B, Min D, Sohn K. A generalized random walk with restart and its application in depth up-sampling and interactive segmentation. IEEE Trans Image Process. 2013; 22(7):2574–88. https://doi.org/10.1109/TIP.2013.2253479 PMID: 23529090

**27.** 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.

**28.** Gormen TH, Leiserson CE, Rivest RL, Stein C, editors. Introduction to algorithms: MIT press Cambridge, MA; 1990.

**29.** Zhu L, Zhang YH, Su F, Chen L, Huang T, Cai YD. A Shortest-Path-Based Method for the Analysis and Prediction of Fruit-Related Genes in Arabidopsis thaliana. PLoS One. 2016; 11(7):e0159519. https://doi.org/10.1371/journal.pone.0159519 PMID: 27434024

**30.** 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(Database issue):D808–15. Epub 2012/12/04. https://doi.org/10.1093/nar/gks1094 PMID: 23203871

**31.** Gene Ontology Consortium: going forward. Nucleic Acids Res. 2015; 43(Database issue):D1049–56. Epub 2014/11/28. https://doi.org/10.1093/nar/gku1179 PMID: 25428369

**32.** Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Research. 2000; 28(1):27–30. PMID: 10592173

**33.** 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. https://doi.org/10.1093/nar/gkm908 PMID: 18194960

**34.** Wei L, Xing P, Zeng J, Chen J, Su R, Guo F. Improved prediction of protein-protein interactions using novel negative samples, features, and an ensemble classifier. Artif Intell Med. 2017.

**35.** Jansen R, Yu H, Greenbaum D, Kluger Y, Krogan NJ, Chung S, et al. A Bayesian networks approach for predicting protein-protein interactions from genomic data. Science. 2003; 302(5644):449–53. https://doi.org/10.1126/science.1087361 PMID: 14564010

**36.** Aloy P, Russell RB. InterPreTS: protein interaction prediction through tertiary structure. Bioinformatics. 2003; 19(1):161–2. PMID: 12499311

**37.** Marcotte EM, Pellegrini M, Ng HL, Rice DW, Yeates TO, Eisenberg D. Detecting protein function and protein-protein interactions from genome sequences. Science. 1999; 285(5428):751–3. PMID: 10427000

**38.** 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. https://doi.org/10.1007/s00438-015-1157-9 PMID: 26728152

**39.** 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. https://doi.org/10.1371/journal.pone.0014556 PMID: 21283518

**40.** Huang G, Chu C, Huang T, Kong X, Zhang Y, Zhang N, et al. Exploring Mouse Protein Function via Multiple Approaches. PLoS One. 2016; 11(11):e0166580. https://doi.org/10.1371/journal.pone.0166580 PMID: 27846315

**41.** 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: a journal of computational molecular cell biology. 2015; 22(1):63–71. Epub 2014/09/24.

**42.** Li Z, An L, Li H, Wang S, Zhou Y, Yuan F, et al. Identifying novel genes and chemicals related to nasopharyngeal cancer in a heterogeneous network. Sci Rep. 2016; 6:25515. Epub 2016/05/07. https://doi.org/10.1038/srep25515 PMID: 27149165

**43.** Wang S, Huang G, Hu Q, Zou Q. A network-based method for the identification of putative genes related to infertility. Biochim Biophys Acta. 2016; 1860(11 Pt B):2716–24. Epub 2016/04/23. https://doi.org/10.1016/j.bbagen.2016.04.010 PMID: 27102279

**44.** Chen L, Wang B, Wang S, Yang J, Hu J, Xie Z, et al. OPMSP: A computational method integrating protein interaction and sequence information for the identification of novel putative oncogenes. Protein Pept Lett. 2016; 23(12):1081–94. PMID: 27774893

**45.** Chen L, Yang J, Huang T, Kong X, Lu L, Cai Y-D. Mining for novel tumor suppressor genes using a shortest path approach. Journal of Biomolecular Structure and Dynamics. 2016; 34(3):664–75. https://doi.org/10.1080/07391102.2015.1042915 PMID: 26209080

**46.** Kohler S, Bauer S, Horn D, Robinson PN. Walking the interactome for prioritization of candidate disease genes. The Amerian Journal of Human Genetics. 2008; 82(4):949–58.

**47.** Tolga C, Çamoğlu O, Singh AK. Analysis of protein-protein interaction networks using random walks. Proceedings of the 5th international workshop on Bioinformatics; Chicago, Illinois. 1134042: ACM; 2005. p. 61–8.

48. Yuan F, Zhang YH, Wan S, Wang S, Kong XY. Mining for Candidate Genes Related to Pancreatic Cancer Using Protein-Protein Interactions and a Shortest Path Approach. Biomed Res Int. 2015; 2015:623121. Epub 2015/11/28. https://doi.org/10.1155/2015/623121 PMID: 26613085

49. Wang B, Yuan F, Kong X, Hu LD, Cai YD. Identifying Novel Candidate Genes Related to Apoptosis from a Protein-Protein Interaction Network. Comput Math Methods Med. 2015; 2015:715639. Epub 2015/11/07. https://doi.org/10.1155/2015/715639 PMID: 26543496

50. Chen L, Chu C, Lu J, Kong X, Huang T, Cai Y-D. A computational method for the identification of new candidate carcinogenic and non-carcinogenic chemicals. Molecular BioSystems. 2015; 11(9):2541–50. https://doi.org/10.1039/c5mb00276a PMID: 26194467

51. Sharan R, Ulitsky I, Shamir R. Network-based prediction of protein function. Mol Syst Biol. 2007; 3:88. Epub 2007/03/14. https://doi.org/10.1038/msb4100129 PMID: 17353930

52. Ng KL, Ciou JS, Huang CH. Prediction of protein functions based on function-function correlation relations. Comput Biol Med. 2010; 40(3):300–5. https://doi.org/10.1016/j.compbiomed.2010.01.001 PMID: 20089249

53. Yang J, Chen L, Kong X, Huang T, Cai YD. Analysis of tumor suppressor genes based on gene ontology and the KEGG pathway. PLoS One. 2014; 9(9):e107202. Epub 2014/09/11. https://doi.org/10.1371/journal.pone.0107202 PMID: 25207935

54. Zhang J, Xing Z, Ma M, Wang N, Cai YD, Chen L, et al. Gene ontology and KEGG enrichment analyses of genes related to age-related macular degeneration. Biomed Res Int. 2014; 2014:450386. Epub 2014/08/29. https://doi.org/10.1155/2014/450386 PMID: 25165703

55. Chen L, Zhang YH, Zheng M, Huang T, Cai YD. Identification of compound-protein interactions through the analysis of gene ontology, KEGG enrichment for proteins and molecular fragments of compounds. Molecular genetics and genomics: MGG. 2016; 291(6):2065–79. Epub 2016/08/18. https://doi.org/10.1007/s00438-016-1240-x PMID: 27530612

56. Chen L, Zhang Y-H, Lu G, Huang T, Cai Y-D. Analysis of cancer-related lncRNAs using gene ontology and KEGG pathways. Artificial Intelligence in Medicine. 2017; 76:27–36. https://doi.org/10.1016/j.artmed.2017.02.001 PMID: 28363286

57. Zhang T, Jiang M, Chen L, Niu B, Cai Y. Prediction of Gene Phenotypes Based on GO and KEGG Pathway Enrichment Scores. BioMed Research International. 2013; 2013:7.

58. Callis J, Carpenter T, Sun CW, Vierstra RD. Structure and evolution of genes encoding polyubiquitin and ubiquitin-like proteins in Arabidopsis thaliana ecotype Columbia. Genetics. 1995; 139(2):921–39. PMID: 7713442

59. Kawasaki T, Nam J, Boyes DC, Holt BF 3rd, Hubert DA, Wiig A, et al. A duplicated pair of Arabidopsis RING-finger E3 ligases contribute to the RPM1- and RPS2-mediated hypersensitive response. Plant J. 2005; 44(2):258–70. https://doi.org/10.1111/j.1365-313X.2005.02525.x PMID: 16212605

60. Lahaie N, Kralikova M, Prezeau L, Blahos J, Bouvier M. Post-endocytotic Deubiquitination and Degradation of the Metabotropic -Aminobutyric Acid Receptor by the Ubiquitin-specific Protease 14. Journal of Biological Chemistry. 2016; 291(13):7156–70. https://doi.org/10.1074/jbc.M115.686907 PMID: 26817839

61. Quatrini L, Molfetta R, Zitti B, Peruzzi G, Fionda C, Capuano C, et al. Ubiquitin-dependent endocytosis of NKG2D-DAP10 receptor complexes activates signaling and functions in human NK cells. Science Signaling. 2015; 8(400):ra108. ra108. https://doi.org/10.1126/scisignal.aab2724 PMID: 26508790

62. Rizzo AA, Salerno PE, Bezsonova I, Korzhnev DM. NMR Structure of the Human Rad18 Zinc Finger in Complex with Ubiquitin Defines a Class of UBZ Domains in Proteins Linked to the DNA Damage Response. Biochemistry. 2014; 53(37):5895–906. https://doi.org/10.1021/bi500823h PMID: 25162118

63. El-Sharkawy I, Sherif S, El Kayal W, Jones B, Li Z, Sullivan AJ, et al. Overexpression of plum auxin receptor PslTIR1 in tomato alters plant growth, fruit development and fruit shelf-life characteristics. BMC Plant Biol. 2016; 16:56. https://doi.org/10.1186/s12870-016-0746-z PMID: 26927309

64. Liu JH, Zhang J, Jia CH, Zhang JB, Wang JS, Yang ZX, et al. The interaction of banana MADS-box protein MuMADS1 and ubiquitin-activating enzyme E-MuUBA in post-harvest banana fruit. Plant Cell Rep. 2013; 32(1):129–37. https://doi.org/10.1007/s00299-012-1347-4 PMID: 23007689

65. Bueso E, Ibanez C, Sayas E, Munoz-Bertomeu J, Gonzalez-Guzman M, Rodriguez PL, et al. A forward genetic approach in Arabidopsis thaliana identifies a RING-type ubiquitin ligase as a novel determinant of seed longevity. Plant Sci. 2014; 215:110–6. https://doi.org/10.1016/j.plantsci.2013.11.004 PMID: 24388521

66. Yoshida H, Nagata M, Saito K, Wang KL, Ecker JR. Arabidopsis ETO1 specifically interacts with and negatively regulates type 2 1-aminocyclopropane-1-carboxylate synthases. BMC Plant Biol. 2005; 5:14. https://doi.org/10.1186/1471-2229-5-14 PMID: 16091151

67. Lemus L, Goder V. Regulation of Endoplasmic Reticulum-Associated Protein Degradation (ERAD) by Ubiquitin. Cells. 2014; 3(3):824–47. https://doi.org/10.3390/cells3030824 PMID: 25100021

68. Li M, Tang D, Wang K, Wu X, Lu L, Yu H, et al. Mutations in the F-box gene LARGER PANICLE improve the panicle architecture and enhance the grain yield in rice. Plant Biotechnol J. 2011; 9(9):1002–13. https://doi.org/10.1111/j.1467-7652.2011.00610.x PMID: 21447055

69. Liu JX, Howell SH. Managing the protein folding demands in the endoplasmic reticulum of plants. The New phytologist. 2016; 211(2):418–28. https://doi.org/10.1111/nph.13915 PMID: 26990454

70. Sun CW, Griffen S, Callis J. A model for the evolution of polyubiquitin genes from the study of Arabidopsis thaliana ecotypes. Plant Mol Biol. 1997; 34(5):745–58. PMID: 9278165

71. Roth U, von Roepenack-Lahaye E, Clemens S. Proteome changes in Arabidopsis thaliana roots upon exposure to Cd2+. Journal of Experimental Botany. 2006; 57(15):4003–13. https://doi.org/10.1093/jxb/erl170 PMID: 17075075

72. Liu X, Hu XM, Jin LF, Shi CY, Liu YZ, Peng SA. Identification and transcript analysis of two glutamate decarboxylase genes, CsGAD1 and CsGAD2, reveal the strong relationship between CsGAD1 and citrate utilization in citrus fruit. Molecular Biology Reports. 2014; 41(9):6253–62. https://doi.org/10.1007/s11033-014-3506-x PMID: 24976574

73. Wang DK, Liu HQ, Li SJ, Zhai GW, Shao JF, Tao YZ. Characterization and molecular cloning of a serine hydroxymethyltransferase 1 (OsSHM1) in rice. J Integr Plant Biol. 2015; 57(9):745–56. https://doi.org/10.1111/jipb.12336 PMID: 25641188

74. Jander G, Norris SR, Joshi V, Fraga M, Rugg A, Yu SX, et al. Application of a high-throughput HPLC-MS/MS assay to Arabidopsis mutant screening; evidence that threonine aldolase plays a role in seed nutritional quality. Plant J. 2004; 39(3):465–75. https://doi.org/10.1111/j.1365-313X.2004.02140.x PMID: 15255874

75. Zhang Y, Sun KH, Sandoval FJ, Santiago K, Roje S. One-carbon metabolism in plants: characterization of a plastid serine hydroxymethyltransferase. Biochemical Journal. 2010; 430:97–105. https://doi.org/10.1042/BJ20100566 PMID: 20518745

76. Matella NJ, Braddock RJ, Gregory JF, Goodrich RM. Capillary electrophoresis and high-performance liquid chromatography determination of polyglutamyl 5-methyltetrahydrofolate forms in citrus products. Journal of agricultural and food chemistry. 2005; 53(6):2268–74. https://doi.org/10.1021/jf048285c PMID: 15769167

77. Meng HY, Jiang L, Xu BS, Guo WZ, Li JL, Zhu XQ, et al. Arabidopsis Plastidial Folylpolyglutamate Synthetase Is Required for Seed Reserve Accumulation and Seedling Establishment in Darkness. Plos One. 2014; 9(7):e101905. e101905. https://doi.org/10.1371/journal.pone.0101905 PMID: 25000295

78. Wei ZY, Sun KH, Sandoval FJ, Cross JM, Gordon C, Kang C, et al. Folate polyglutamylation eliminates dependence of activity on enzyme concentration in mitochondrial serine hydroxymethyltransferases from Arabidopsis thaliana. Archives of Biochemistry and Biophysics. 2013; 536(1):87–96. https://doi.org/10.1016/j.abb.2013.06.004 PMID: 23800877

79. Barkla BJ, Vera-Estrella R, Miranda-Vergara MC, Pantoja O. Quantitative proteomics of heavy metal exposure in Arabidopsis thaliana reveals alterations in one-carbon metabolism enzymes upon exposure to zinc. Journal of proteomics. 2014; 111:128–38. https://doi.org/10.1016/j.jprot.2014.03.006 PMID: 24642212

80. Lugassi N, Kelly G, Fidel L, Yaniv Y, Attia Z, Levi A, et al. Expression of Arabidopsis Hexokinase in Citrus Guard Cells Controls Stomatal Aperture and Reduces Transpiration. Frontiers in Plant Science. 2015; 6:1114. 1114. https://doi.org/10.3389/fpls.2015.01114 PMID: 26734024

81. Chen JG, Jones AM. AtRGS1 function in Arabidopsis thaliana. Regulators of G-Protein Signaling, Part A. 2004; 389:338–50.

82. Dyson BC, Webster RE, Johnson GN. GPT2: a glucose 6-phosphate/phosphate translocator with a novel role in the regulation of sugar signalling during seedling development. Ann Bot-London. 2014; 113 (4):643–52.

83. Bueso E, Munoz-Bertomeu J, Campos F, Brunaud V, Martinez L, Sayas E, et al. ARABIDOPSIS THALIANA HOMEOBOX25 Uncovers a Role for Gibberellins in Seed Longevity. Plant Physiol. 2014; 164(2):999–1010. https://doi.org/10.1104/pp.113.232223 PMID: 24335333

84. Pracharoenwattana I, Cornah JE, Smith SM. Arabidopsis peroxisomal citrate synthase is required for fatty acid respiration and seed germination. Plant Cell. 2005; 17(7):2037–48. https://doi.org/10.1105/tpc.105.031856 PMID: 15923350

85. Yang XY, Chen ZW, Xu T, Qu Z, Pan XD, Qin XH, et al. Arabidopsis Kinesin KP1 Specifically Interacts with VDAC3, a Mitochondrial Protein, and Regulates Respiration during Seed Germination at Low Temperature. Plant Cell. 2011; 23(3):1093–106. https://doi.org/10.1105/tpc.110.082420 PMID: 21406623

86. Anoman AD, Munoz-Bertomeu J, Rosa-Tellez S, Flores-Tornero M, Serrano R, Bueso E, et al. Plastidial Glycolytic Glyceraldehyde-3-Phosphate Dehydrogenase Is an Important Determinant in the Carbon and Nitrogen Metabolism of Heterotrophic Cells in Arabidopsis. Plant Physiol. 2015; 169(3):1619–37. https://doi.org/10.1104/pp.15.00696 PMID: 26134167

87. Andriotis VME, Kruger NJ, Pike MJ, Smith AM. Plastidial glycolysis in developing Arabidopsis embryos. New Phytologist. 2010; 185(3):649–62. https://doi.org/10.1111/j.1469-8137.2009.03113.x PMID: 20002588

88. Wei L, Tang J, Zou Q. Local-DPP: An improved DNA-binding protein prediction method by exploring local evolutionary information. Inform Sciences. 2017; 384:135–44.