

Article

Network Analysis of Local Gene Regulators in *Arabidopsis thaliana* under Spaceflight Stress

Vidya Manian ^{1,2,*}, Harshini Gangapuram ², Jairo Orozco ¹, Heeralal Janwa ³ and Carlos Agrinsoni ³

¹ Department of Electrical and Computer Engineering, University of Puerto Rico, Mayaguez, PR 00681-9000, USA; jairo.orozco@upr.edu

² Department of Bioengineering, University of Puerto Rico, Mayaguez, PR 00681-9000, USA; harshini.gangapuram@upr.edu

³ Department of Mathematics, University of Puerto Rico, Rio Piedras, PR 00925-2537, USA; heeralal.janwa@upr.edu (H.J.); carlos.agrinsoni@upr.edu (C.A.)

* Correspondence: vidya.manian@upr.edu

Abstract: Spaceflight microgravity affects normal plant growth in several ways. The transcriptional dataset of the plant model organism *Arabidopsis thaliana* grown in the international space station is mined using graph-theoretic network analysis approaches to identify significant gene transcriptions in microgravity essential for the plant's survival and growth in altered environments. The photosynthesis process is critical for the survival of the plants in spaceflight under different environmentally stressful conditions such as lower levels of gravity, lesser oxygen availability, low atmospheric pressure, and the presence of cosmic radiation. Lasso regression method is used for gene regulatory network inferencing from gene expressions of four different ecotypes of *Arabidopsis* in spaceflight microgravity related to the photosynthetic process. The individual behavior of hub-genes and stress response genes in the photosynthetic process and their impact on the whole network is analyzed. Logistic regression on centrality measures computed from the networks, including average shortest path, betweenness centrality, closeness centrality, and eccentricity, and the HITS algorithm is used to rank genes and identify interactor or target genes from the networks. Through the hub and authority gene interactions, several biological processes associated with photosynthesis and carbon fixation genes are identified. The altered conditions in spaceflight have made all the ecotypes of *Arabidopsis* sensitive to dehydration-and-salt stress. The oxidative and heat-shock stress-response genes regulate the photosynthesis genes that are involved in the oxidation-reduction process in spaceflight microgravity, enabling the plant to adapt successfully to the spaceflight environment.

Keywords: gene expression; network analysis; logistic regression; HITS algorithm; photosynthesis; carbon fixation; biological processes; spaceflight microgravity; *Arabidopsis thaliana*



Citation: Manian, V.; Gangapuram, H.; Orozco, J.; Janwa, H.; Agrinsoni, C. Network Analysis of Local Gene Regulators in *Arabidopsis thaliana* under Spaceflight Stress. *Computers* **2021**, *10*, 18. <https://doi.org/10.3390/computers10020018>

Academic Editor: Lucia Maddalena
Received: 28 December 2020
Accepted: 22 January 2021
Published: 28 January 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Arabidopsis thaliana (*Arabidopsis*), a member of the Brassicaceae or mustard family, is a small photosynthetic plant that requires only light, air, water, and few minerals for its survival. It occupies minimal space and can be quickly grown in an indoor growth chamber. *Arabidopsis* is an excellent model plant because of its small genomic size, low DNA content, and its genetic manipulability [1]. *Arabidopsis* genome contains about 25,500 genes containing approximately 35% unique genes [2].

Arabidopsis has several ecotypes/natural variants, and these variations can sometimes be visible in physical traits when they are grown under different environmental stressors on the ground. Besides exploring natural genetic variants of *Arabidopsis* under different environmental stressors, the genetic factors causing physiological variations are mostly unknown [3]. The genetic basis of responses of different ecotypes/natural variants of *Arabidopsis* under different environmental stressors such as hypoxia, light, dark, salt, drought,

heat shock on the ground depending on different geographical conditions have been investigated [4]. These investigations include but are not limited to studying environmental stressors such as differential drought responses [5], cold stress responses [6], response to salt stress [7], heat stress responses [6], enhanced stress tolerance [8], and effects of chronic ozone exposure [9] on ecotypes of *Arabidopsis*.

Investigations on genetic variations in the photosynthetic process in different *Arabidopsis* ecotypes in spaceflight can reveal essential cues that are useful for its growth in stressful environments such as in space stations and on Moon or Mars. As sufficient transcriptomic data analyses for different ecotypes of *Arabidopsis* are available on the ground, *Arabidopsis* is grown by scientists in the International Space Station (ISS) to compare its transcriptomic stress responses in spaceflight microgravity with that of ground [10]. Microgravity, elevated levels of solar energy, and galactic cosmic radiation also influence the plants grown in the ISS [11].

Photosynthesis is one of the critical biological processes on Earth that uses sunlight for carbon dioxide (CO₂) fixation in plants. Photosynthesis is responsible for the growth and other energy-dependent metabolic pathways in plants. In the oxygenic photosynthetic process, the light absorbed by the plant is stored as chemical energy. The absorption of light is done in pigment-containing holoprotein complexes in the thylakoid membrane as photosystem I (PSI) and photosystem II (PSII). Upon absorption of light in PSII, then the light energy is used to transfer electrons from water (H₂O) molecule to CO₂ to produce carbohydrates. In this process, the water molecule becomes oxidized, and oxygen (O₂) is released. Therefore, in oxygenic photosynthesis, light energy leads to the carbon fixation pathway in plants. The genes that are responsible for photosynthesis in *Arabidopsis* are discussed in [12].

When exposed to excessive doses of light, the plants receive more light energy than required for photosynthesis, which results in the production of harmful reactive oxygen species in the cells that causes irreversible photo-oxidative damage. The photo-oxidative damage inhibits the PSII process, which may lead to loss of productivity of oxygen in plants [13]. Hence, it is crucial to study the behavior of the photosynthetic genes in spaceflight microgravity when compared to ground control. In [11], the authors discuss how the transcripts of heat-shock proteins are upregulated, and the transcripts of peroxidase are downregulated in the spaceflight environment. The gene classes mentioned in [11] are related to oxidative stress and hypoxia. In this paper, we analyze the GeneLab dataset (GLDS-37) collected from the experiment carried out in spaceflight [11], for spaceflight stress response (see Supplementary Materials).

Choosing the best optimization method to process the transcriptomic data is always challenging for investigators. Many optimization algorithms are available in recent years for finding the correlation between the genes in the gene expression data. Network analysis has considerably helped in the analysis of transcriptomic data as they serve as a “blueprint” to study the molecular interactions. One can understand the regulation of genes by network analysis [14]. Moreover, partial Gene Regulatory Networks (GRN) responsible for a biological process can be retrieved and compared for the regulation of genes in different stress conditions. Integrating GRN with enrichment analysis tools such as Gene Set Enrichment Analysis (GSEA) [15] and Kyoto Encyclopedia of Genes and Genomes (KEGG) [16] helps investigators to examine the involvement of a set of genes and their regulation under different environmental stressors in specific biological processes. The underlying mechanisms of the GRN are revealed through topological and algebraic analysis of the GRN [17,18]. The individual behavior of the genes or a set of genes can be extracted from the GRN, enabling us to understand the impact of these genes on the whole GRN.

2. Materials and Methods

In this paper, we have analyzed the transcriptomic data of different ecotypes of *Arabidopsis* in spaceflight microgravity and compared them to ground control by generating gene regulatory networks. The photosynthesis and carbon fixation genes are extracted by

performing GSEA and KEGG pathway analysis. Dimensionality reduction of the expression values of these genes is made using Principal Components Analysis (PCA) to eliminate noise and extract independent features of the genes. The first three components of PCA is selected because it captures 99% of the variance of the gene expression flight and ground datasets. The Lasso regression is performed on these three components of PCA to find correlation among the gene expression values. The adjacency matrix is constructed, the source-target lists are made, and the GRN is visualized in Cytoscape. Using the software, including Python, SAGE, we carry out Pearson correlation GRN inferencing, logistic regression, and HITS ranking of genes and topological analysis in a novel way to identify hub and authority genes. The topological analysis involves computations of various centrality measures. Figure 1 shows a general flow diagram for GRN construction and network analysis methods that can be applied to omics data.

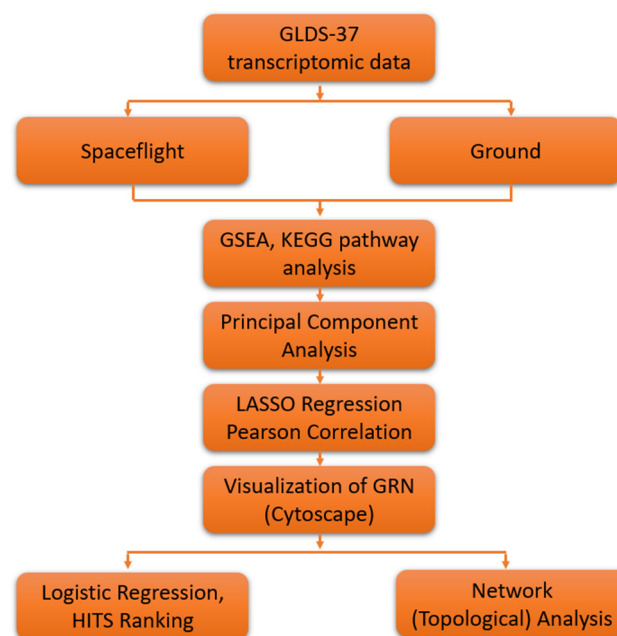


Figure 1. Flow chart for GRN construction and network analysis.

2.1. Genelab *Arabidopsis thaliana* Dataset

GLDS-37 presents the transcriptomic data of four different ecotypes of *Arabidopsis* in spaceflight microgravity and on the ground: Col-0, Cvi, LER, and WS. The seeds of the mentioned ecotypes were germinated in orbit and grown for eight days. The same environmental stressors are maintained on the ground to observe the behavior of the plants. Later, RNAseq was performed to catalog the differential expression of the ecotypes. The primary purpose of this study is to analyze the stress response mechanisms of these different ecotypes of *Arabidopsis* under oxidative stress and hypoxia. The datasets, their description, and relevant details are available at <https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-37/>.

2.2. Gsea and Kegg Pathway Analysis

The genes responsible for photosynthesis and carbon fixation are retrieved by performing GSEA and KEGG pathway analysis. GSEA is a tool that can identify the group of genes from the gene expression data that are responsible for a shared biological process [15]. GSEA (<https://www.gsea-msigdb.org/gsea/index.jsp>) is performed on datasets corresponding to different ecotypes of *Arabidopsis* with the *ShinyGO* tool (<http://bioinformatics.sdstate.edu/go/>). With *KEGGmapper* (<https://www.genome.jp/kegg/mapper.html>), one can convert protein-coding genes in a genome to KEGG molecular networks describing a

pathway relating the genes involved in different cellular functions and high-level processes (for example, photosynthesis) [19].

2.3. Principal Component Analysis

The transcriptomic data have expression values of thousands of genes from different experimental conditions. Hence, it becomes difficult for investigators to determine whether the time series data available is for different states of gene expression or just a measurement for similar states obtained by different mechanisms. PCA is an unsupervised dimensionality reduction algorithm used globally to analyze -omics data [20]. PCA is applied to find the core group of independent features that are available in -omics data [21]. We have done dimensionality reduction of the GLDS-37 dataset using PCA, and subsequently, the first three components of the PCA are given as input to the Lasso regression algorithm to compute the correlation between the genes.

2.4. Lasso Regression

GRN reconstruction is a common problem in computational biology for which various methods have been proposed, systemically assessed, and reviewed [22]. GRN inferencing on transcriptomic data based on the application of a similarity measure on the dataset results in a similarity matrix. This similarity matrix undergoes multiple hypothesis tests to determine the statistical significance between the genes. As a result, we obtain a sparse matrix that includes both the direct and indirect relationships between the genes [23]. A powerful tool is necessary to identify the direct correlation between the genes to infer GRN. The regression-based models such as Lasso can extract one-to-many relationships between the genes, according to the corresponding transcriptomic data. Lasso is a traditional regularized regression method used to infer GRN with accurate results [24].

The R package SILGGM [25] is used to infer the GRN from the GeneLab transcriptomic data of *Arabidopsis*. The authors have developed two main approaches for estimating the conditional dependence of genes: (i) graphical Lasso, and (ii) a penalized regression, based on neighborhood approach. We have used a scaled Lasso algorithm that has precise conditional dependence estimators that take the first three principal components of the transcriptomic data as inputs. A GRN is a causal relationship between a transcription factor (hub-gene) and a target (interactor or authority) gene. Here, we did not give separate transcription factors to the algorithm. The same set of genes are considered as both transcription factors and target genes, and the statistical inference method calculates the correlation between the genes—the GRNs are visualized in Cytoscape with the network files obtained from scaled Lasso.

2.5. Logistic Regression Based Gene Ranking

Logistic regression, also called a logit model, is a statistical method for analyzing a dataset in which there are one or more independent variables that determine an outcome. The goal of this method is to find the best fitting model to describe the relationship between a dichotomous characteristic of interest (dependent variable; response or outcome variable) and the explanatory variable. It is used to model the log-odds of a gene belonging to a specific category as a linear function of the statistical significance x :

$$\log\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta x \quad (1)$$

where α is the intercept, β is the slope, both α and β are estimated from the data. Most likely enriched gene sets will be identified based on the p -value or based on the odds ratio if a ranking independent of category size is desired [26]. The logistic regression method is an extension of the χ^2 -test and has higher statistical power than other methods because the important values do not depend on a threshold.

2.6. Network Analysis

Network measures such as the average shortest path length, betweenness centrality, closeness centrality, and eccentricity of the hub-genes and stress response genes present in each GRN are calculated. Fold-change analysis is used to determine whether the hub-genes and stress response genes are upregulated or downregulated in spaceflight microgravity when compared to ground control.

2.6.1. Fold-Change

Fold-change is often used to analyze the expression level of genes in microarray and RNA-Seq experiments. Fold change is a \log_2 transformed ratio of the gene expression values from experiment vs. control [27]. Fold change can be calculated using Equation (2)

$$\log_{2FC} = \log_2 \left(\frac{A}{B} \right) \quad (2)$$

In Equation (2), “A” represents the expression value of a gene in the experiment, and ‘B’ represents the expression value of the same gene in a control environment. The analysis of fold change shows whether the given gene is upregulated or downregulated in spaceflight microgravity when compared to its expression in ground control.

2.6.2. Graphs and Networks

Let $G = (V, E)$ be a graph, where V is the set of vertices and E is the set of edges. Let p_k be the degree distribution and $\langle k \rangle$ be the average degree of the graph G .

Outdegree: The outdegree distribution of a given node (gene) in the network determines whether the node targets (or directs) other nodes [28]. We have calculated the outdegree distributions of hub-genes and stress response genes in the GRN to determine whether the hub genes are regulating other genes.

Average shortest path length: The average shortest path in network topology is defined as the average minimum distance that a node can take to reach all other possible nodes (targets) in the network [29]. The average shortest path length is calculated using Equation (3) [28]:

$$a = \sum_{s,t \in V} \frac{d(s,t)}{n(n-1)} \quad (3)$$

Here, V is the total number of nodes in the network, and $d(s, t)$ is the minimum distance from the node (s) to the node (t). “ n ” denotes the total number of nodes. The average shortest path length is meant to be small for the networks to be small world. Here, one can calculate the average shortest path length of all nodes to evaluate the efficiency of a network or determine the individual average shortest path length of each node. We have measured the average shortest path length of each node to determine the efficiency of each hub-gene and stress response gene in the GRN. We define the average shortest path length for any given node s as:

$$a(s) = \sum_{t \in V} \frac{d(s,t)}{(n-1)} \quad (4)$$

Betweenness centrality: Betweenness centrality is a measure that shows how frequently a given node appears on the shortest paths of other nodes [30]. It acts as a bridge for other nodes to be connected by providing the shortest path. The higher betweenness centrality means that the given node appears very frequently in the shortest paths of a greater number of other nodes. Betweenness centrality can be calculated using Equation (5) [30]:

$$g(\theta) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(\theta)}{\sigma_{st}} \quad (5)$$

The total number of shortest paths from the node (s) to node (t) is represented by σ_{st} . The total number of paths that pass through the node (v) is represented by $\sigma_{st}(\theta)$. The GRN's are directed networks. Therefore, betweenness values are normalized by $\{1/((n-1)(n-2))\}$, where n is the number of nodes in the network.

Closeness centrality: The average shortest path length of a given node and all other nodes in the network are measured by Closeness Centrality [31]. The node that has a higher closeness centrality is the most central node that is connected to the maximum number of nodes in the network.

$$C(x) = \sum_y \frac{(n-1)}{d(x,y)} \quad (6)$$

Closeness centrality is calculated using Equation (6) [31]. Here, n is the total number of shortest paths going through a given node, y is the given node, and x is the node that passes through the node y .

Eccentricity: Eccentricity is a measure of the maximum distance by which a node can be connected to another node [32]. Eccentricity shows how one node is indirectly connected to other nodes in the network in the path to its target node. A higher eccentricity implies that the node has the greatest influence on the network compared to other nodes.

Clustering Coefficient: Clustering coefficient measures how the adjacent vertices connect to each other. Given a vertex v_i we define the clustering coefficient as

$$C_i = \frac{2L_i}{k_i(k_i-1)} \quad (7)$$

where k_i is the degree of vertex v_i and L_i is the number of edges between the adjacent vertices of v_i [33]. The average clustering coefficient is defined as follows:

$$\langle C \rangle = \frac{1}{|V|} \sum_{i=1}^{|V|} C_i. \quad (8)$$

The average clustering coefficient can be interpreted as the probability that two adjacent vertices of a randomly selected vertex are connected to each other [34].

Assortativity coefficient: The assortativity coefficient measures the Pearson correlation coefficient between pairs of adjacent vertices. The assortativity coefficient r is given by

$$r = \frac{\sum_{j,k} jk (e_{jk} - q_j q_k)}{\sigma_q^2} \quad (9)$$

where $q_k = \frac{|V| p_k}{\langle k \rangle}$ is the distribution of the remaining degree and e_{jk} is the probability of finding vertices with degrees j and k as the two ends of a randomly selected edge [35].

HITS algorithm for detecting hubs and authority genes: Our GRN (similar to other complex networks) follow preferential attachment models, which are scale-free with a degree distribution that follows an exponential law. Unlike the random graph model, these networks have nodes with large degrees, called hubs. Classically, with 0–1 (nonconnection-connection) networks, just the degree distribution is used in the identification of such hubs. A much more sophisticated algorithm is proposed by Kleinberg [36] called the Hypertext induced topic search (HITS) algorithm. Originally it was meant for the networks such as the Internet. We use it now to study our GRN (see also [37,38]). Most of these applications use PageRank to reveal localized information about the graph based on some form of external data. We apply this algorithm in our setting for the weighted and directed networks for the transcription factors-target gene networks and co-expression networks.

In the weighted GRN setting, the traditional simplistic method of detecting hub genes would not yield meaningful information. Our approach uses, iteratively the weighted HITS algorithm in a novel way as follows. At the k th iteration, let h_k (respectively a_k) be whose i th entry $h(i)$ (respectively $a(i)$) be the hub weight (authority weight) assigned to node i .

One initially assigns uniform distribution on the nodes. Let $h_k := \sum_u a_{k-1}(j)$, the sum being over authority nodes j pointed to by i . And similarly, the authority node weights are computed by $a_k \sum_j h_{k-1}$. Then, we normalize so that the sum of the weights equal to 1, with normalization factors ψ_k (respectively ϕ_k). In matrix notation: $h_k = \psi_k \phi_{k-1} A A^T a_{k-2}$, and $a_k = \phi_k \psi_{k-1} A^T A a_{k-2}$. These iterations converge to the dominant eigenvector of the real symmetric matrices AA^T (respectively $A^T A$). These give us asymptotically hub and authority weights. In this setup, we have assumed entries in A to be 0 or 1. If there are weights on the edges such as correlation or signal strength w_{ij} , then they are introduced in the sums. Since adjacency is defined for undirected graphs as well, this algorithm will return hubs and authorities weights for such graphs, as well.

We can describe the algorithms as pseudocode as follows, where in place of the 1-norm (sum of the absolute values), we can use any norm (Algorithm 1):

Algorithm 1 pseudocode

```

HITS(A):          # A :=(The adjacency matrix of the weighted network N = (V,E))
Local Variables : n = |V|; e = |E|
h;                # hub rank real vector (in  $\mathbb{R}^n$ )
a;                # authority rank real vector (in  $\mathbb{R}^n$ )
m;                # the number of HITS iterations.
 $h_1 = \frac{1}{n}$ ;    # all entries in h are 1
 $a_1 = \frac{1}{n}$ ;    # all entries in the vector are 1.
 $\psi_1 = \mathbf{1}$ ;
 $\phi_1 = \mathbf{1}$ ;
while k ≤ m do;
begin;
   $h_k = A a_{k-1}$ ;
   $a_k = A^T h_{k-1}$ ;
 $h_k \frac{h_k}{\text{norm}(h_k)}$ ;          # norm(v) = square-root of sum of squares
 $a_k := \frac{a_k}{\text{norm}(a_k)}$ ;
end

```

We have also used different norms in the computations of the normalization factors. The standard norm (square-root of the sum of squares, is proposed by Kleinberg [36]). If the eigenvalues of AA^T (the same as those of $A^T A$) are separated (i.e., the multiplicity of the dominant eigenvalues is 1), then the iterations in the converge in the limit to the corresponding dominant eigenvector of AA^T . In the matrix notation, this is the famous QR decomposition method or the unsymmetric eigenvalue problem (see 7.3.1 of [39]). We used a SAGE implementation package [40].

Since our GRN graph is sparse but highly connected, it converges rapidly with a large number of iterations, yielding hubs and authority genes in this very complex network. These iterations give us asymptotically hub and authority weights. We have implemented a version of this algorithm in the SAGE software and then applied it to our networks to find hub genes and authority genes. Our algorithm gives weighted-hub genes and weighted-authority (target) genes. In complex networks, the HITS algorithm has very high complexity and cannot be applied successfully. The weighted HITS algorithm has yielded important information about biomolecular networks [41–49]. For network topology, we refer to [34], and for the latest on the origin of biomolecular networks, topological, combinatorial, and spectral methods, we refer to [18].

Small World Phenomenon: Biomolecular networks have features that are not captured by the Erdos and Renyi random graph model. As we have seen, random graphs have a low clustering coefficient, and they do not account for the formation of hubs. To rectify some of these shortcomings, the *small world model*, popularly known as the *six degrees of separation model* was introduced as the next level of complexity for a probabilistic model with features that are closer to real world networks [33].

In this model, the graph G of N nodes is constructed as a ring lattice, in which, (i) first, **wire**: that is, connect every node to $K/2$ neighbors on each side and (ii) second, **rewire**: that is, for every edge connecting a particular node, with probability p reconnect it to a randomly selected node. The average number of such edges is $pNK/2$. The first step of the algorithm produces local clustering, while the second dramatically reduces the distance in the network. Unlike random graphs, the clustering coefficient of this network $C = 3(K - 2)/4(K - 1)$ is independent of the system size. Thus, the small world network model displays the small world property and the clustering of real networks, however, it does not capture the emergence of hubby nodes (e.g., p53 in biomolecular networks) (part of one of the eight open problems that we formulate in Section 4 in [18]).

Scale-free Network Models: Most biomolecular networks are hypothesized to have a degree distribution, described as *scale-free*. In a scale free network, the number of nodes n_k of degree k is proportional to a power of the degree, namely, the degree distribution of the nodes follows a *power-law*

$$n_k = k^{-\beta} \quad (10)$$

where $\beta > 1$ is a coefficient characteristic of the network. Unlike in random networks, where the degree of all nodes is centered around a single value—with the probability of finding nodes with much larger (or smaller) degree decaying exponentially, in scale-free networks, there are nodes of large degree with relatively higher probability (fat tail). In other words, since the power law distribution decreases much more slowly than exponentially, for large k (heavy or fat tails), scale-free networks support nodes with the extremely high number of connections called “hubs.” Power law distribution has been observed in many large networks, such as the Internet, the phone-call maps, and other collaboration networks [34]. A caveat to these reports is that inappropriate statistical techniques have often been used to infer power law distributions, and alternative heavy tailed distributions may fit the data better. However, the power law is a useful approximation that allows mechanisms of network growth to be explored, such as preferential attachment, discussed next, while the examination of alternative heavy tailed distributions is set as an open problem.

Preferential Attachment: The original model of preferential attachment was proposed by Barabási–Albert [34]. The scheme consists of a local *growth rule* that leads to a global consequence, namely a power law distribution. The network grows through the addition of new nodes linking to nodes already present in the system. There is a higher probability to preferentially link to a node with a large number of connections. Thus, this rule gives more preferences to those vertices that have larger degrees. For this reason, it is often referred to as the “rich-get-richer” or “Matthew” effect. This can be formulated as a game-theoretic problem originating from information asymmetry and associated Nash equilibrium, discussed in the Open Problems.

With an initial graph G_0 and a fixed probability parameter p , the preferential attachment random graph model $G(p, G_0)$ can be described as follows: at each step the graph G_t is formed by modifying the earlier graph G_{t-1} in two steps—with probability p take a vertex-step; otherwise, take an edge-step:

- Vertex step: Add a new vertex v and an edge $\{u, v\}$ from v to u by randomly and independently choosing u proportional its degree;
- Edge step: Add a new edge $\{r, s\}$ by independently choosing vertices r and s with probability proportional to their degrees.

That is, at each step, we add a vertex with probability p , while for sure, we add an additional edge. If we denote by n_t and e_t the number of vertices and edges respectively at step t , then $e_t = t + 1$ and $n_t = 1 + \sum_{i=1}^t z_i$, where z_i 's are Bernoulli random variables with the probability of success = p . Hence the expected value of nodes is $\langle n_t \rangle = 1 + pt$.

It can be shown that exponentially (as t asymptotically approaches infinity) this process leads to a scale-free network. The degree distribution of $G(p)$ satisfies a power law with the parameter for exponent being $\beta = 2 + \frac{p}{2-p}$. Scale-free networks also exhibit hierarchicity. The local clustering coefficient is proportional to a power of the node degree

$$C(k) \approx k^{-\alpha} \quad (11)$$

where α is called the hierarchy coefficient.

This distribution implies that the low-degree nodes belong to very dense sub-graphs and those sub-graphs are connected to each other through hubs. In other words, it means that the level of clustering is much larger than that in random networks.

Consequently, many of the network properties in a scale-free network are determined by local structures as observed in a relatively small number of highly connected nodes (hubs). A consequence of this scale-free network property is its extreme robustness to failure, which is also displayed by biomolecular networks and their modular structures. Such networks are highly tolerant of random failures (perturbations); however, they remain extremely sensitive to targeted attacks.

3. Results

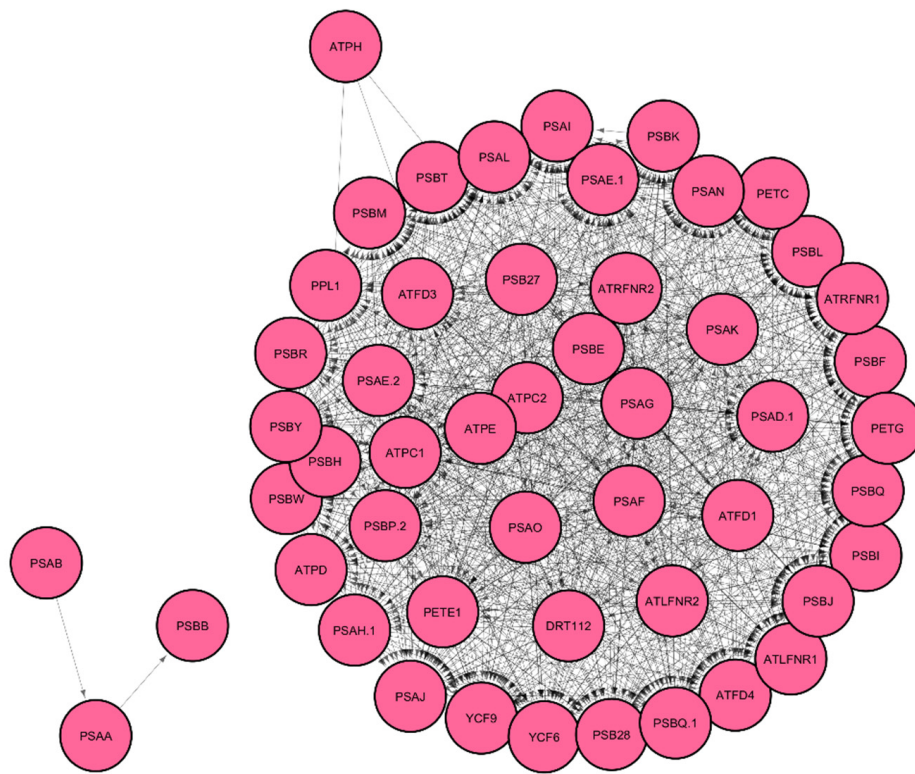
Figure 2 shows the GRN of Col-0 ecotype in spaceflight microgravity and ground control for photosynthesis and carbon fixation biological processes. All the GRN of other ecotypes are constructed in the same manner.

3.1. Identification of Regulatory Hub-Genes in Photosynthesis and Carbon Fixation Grn

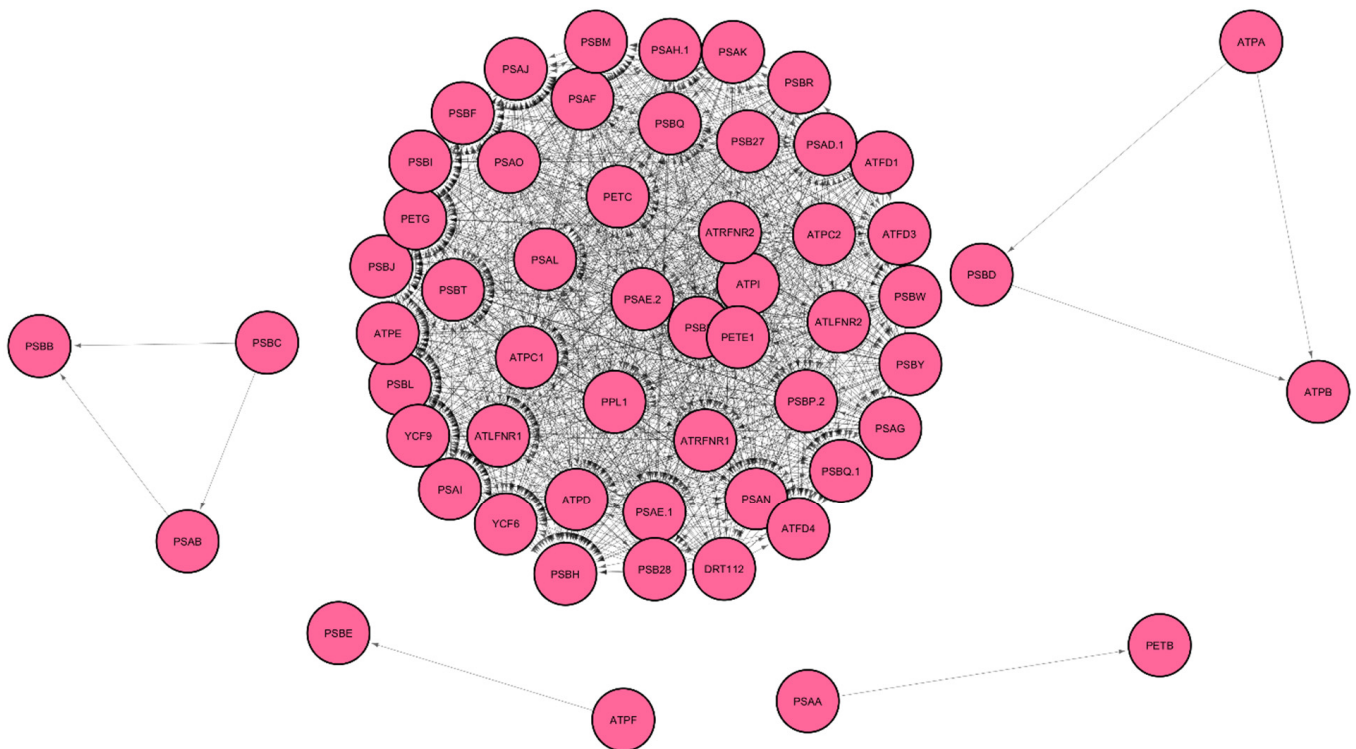
The most significant regulatory genes (hub-genes) that influence the maximum number of genes (maximum outdegree) in photosynthesis and carbon fixation GRN are isolated. The hub-genes act as transcription factors (TFs) that regulate other target genes (nodes) in the GRN.

The common hub-genes of all the ecotypes in spaceflight microgravity and ground control related to photosynthetic GRN are DRT112, ATRFNR2, PSAK, PSB27, ATFD1, PSAF, ATLFNR2, ATPC2, PSAO. The hub-genes in the WS ecotype in spaceflight microgravity are different compared to other GRN. The hub-gene PSBY is seen only in the GRN of WS in spaceflight microgravity. The hub-gene PSAG that is present in other GRN is not present in GRN of WS in spaceflight microgravity. We have done gene ontology analysis (refer to Table 1) to understand the functions of each hub-gene. PSB27, PSAO, ATFD1, and DRT112 are involved in the generation of precursor metabolites and energy. PSAO and PSB27 together perform light-harvesting and reaction mechanisms to light in the photosynthetic process. ATFD1, DRT112, and PASO are the hub-genes that participate in the electron transport chain. The genes ATFD1, ATLFNR2, DRT112, ATRFNR2, and PSAO play a vital role in the oxidation-reduction process.

The common hub-genes of all the ecotypes in spaceflight microgravity and ground control in carbon fixation GRN are RSW10, RPI2, AOAT2, ALAAT2, ASP4, and ATPPC4. Apart from the common hub-genes, the carbon fixation GRN are observed to have other hub-genes that are present only in the respective GRN. This shows that the ecotypes have different transcriptional regulators for each ecotype in spaceflight microgravity and ground control. Gene ontology results reveal that the hub-genes RPI2, AOAT2, ALAAT2, ASP4, and ATPPC4 are involved in the carbon metabolism of carbon fixation in the photosynthetic process. The genes that biosynthesize amino acids are RSW10, RPI2, ALAAT2, ASP4. The genes that are involved in metabolic pathways are RSW10, RPI2, AOAT2, ALAAT2, ASP4, and ATPPC4.



(A)



(B)

Figure 2. Cont.

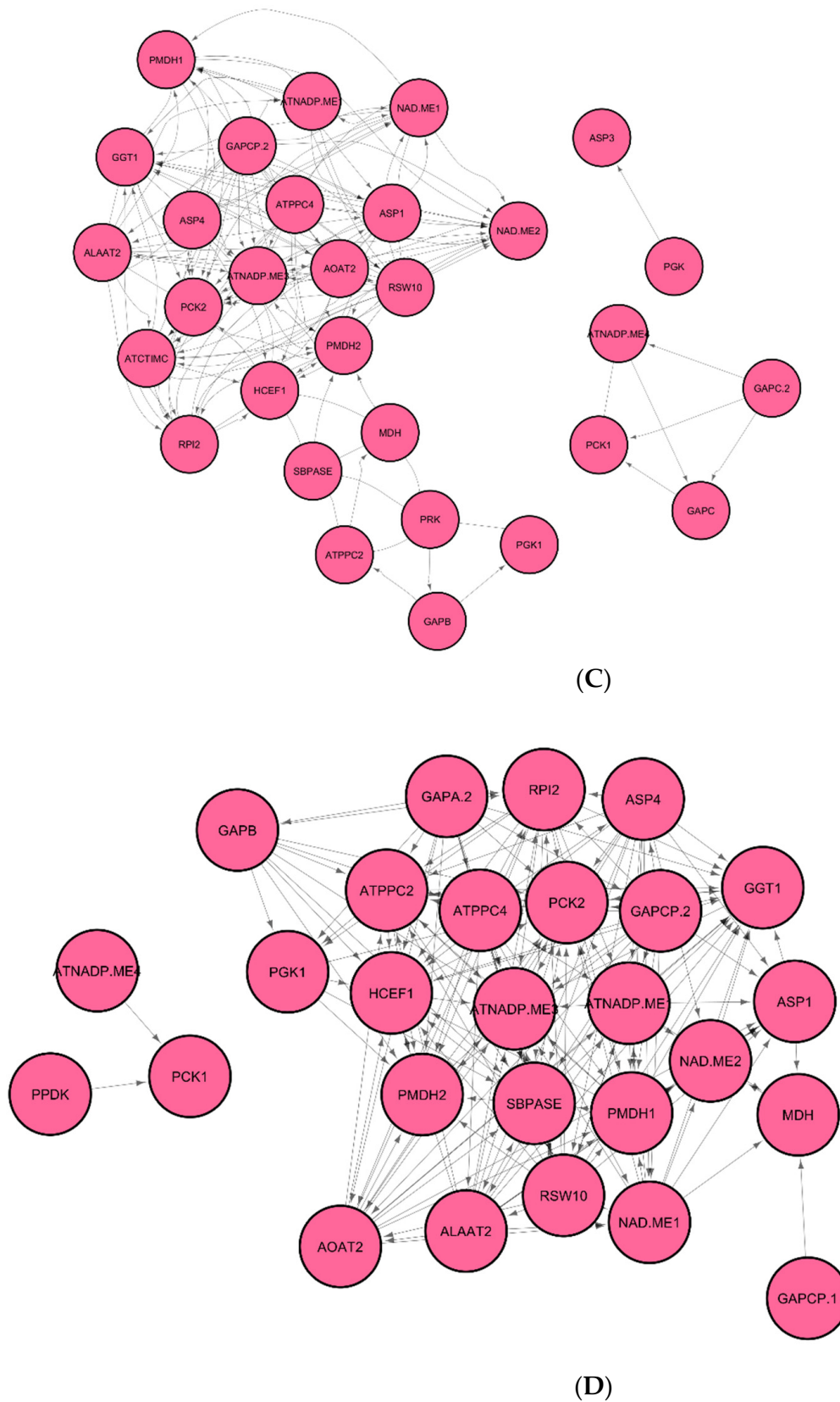


Figure 2. The photosynthetic and carbon fixation GRN of Col-0 ecotype in spaceflight microgravity and ground control. **(A)** Photosynthetic GRN of Col-0 ecotype in spaceflight microgravity. **(B)** Photosynthetic GRN of Col-0 ecotype in ground control. **(C)** Carbon fixation GRN of Col-0 ecotype in spaceflight microgravity. **(D)** Carbon fixation GRN of Col-0 ecotype in ground control.

Table 1. Gene ontology of the hub-genes and stress response genes involved in photosynthesis and carbon fixation.

Hub-Genes and Stress Response Genes	Name	Gene Ontology (GO) Term
PSAO	PSI subunit O	Part of integral component of membrane
PSB27	PSII lipoprotein	Involved in PSII assembly
ATFD1	Ferredoxin 1	Part of chloroplast
ATPC2	ATP synthase subunit A	Involved in ATP synthesis coupled proton transport
DRT112	Plastocyanin	Involved in oxidation-reduction process
ATLFNR2	Ferredoxin—NADP reductase	Part of chloroplast, involved in oxidation-reduction process
ATRFNR2	Ferredoxin-NADP reductase	Involved in oxidation-reduction process
PSAK	PSI reaction center subunit	Part of chloroplast, involved in PSI
PSAF	PSI subunit III	Part of PSI reaction center
RSW10	Ribose-5-phosphate isomerase 1	Enables ribose-5-phosphate isomerase activity
ALAAT2	Alanine_1_2 domain-containing protein	Enables transferase activity
AOAT2	Glutamate-glyoxylate aminotransferase 2	Involved in photorespiration
RPI2	Ribose-5-phosphate isomerase 2	Enables isomerase activity
ATPPC4	Phosphoenolpyruvate carboxylase 4	Involved in carbon fixation
ASP4	Aspartate aminotransferase	Involved in cellular amino acid metabolic process
PETC	Cytochrome b6-f complex iron-sulfur subunit	Involved in oxidation-reduction process
ATPD	ATP synthase subunit beta	Involved in ATP synthesis coupled proton transport
ATLFNR1	Ferredoxin—NADP reductase	Enables oxidoreductase activity
PRK	Parkin-FBXW7-Cul1 ubiquitin ligase complex	Enables phosphoribulokinase activity
GAPB	Gp_dh_N domain-containing protein	Enables nucleotide binding
ASP1	Accessory Sec system protein	Involved in protein transport
PGK1	Phosphoglycerate kinase 1	Enables phosphoglycerate kinase activity
GAPC	Glyceraldehyde-3-phosphate dehydrogenase	Enables oxidoreductase activity
MDH	Malate dehydrogenase 1	Involved in carbohydrate metabolic process
ATCTIMC	Putative triosephosphate isomerase	Involved in glyceraldehyde-3-phosphate biosynthetic process
SBPASE	FBPase domain-containing protein	Involved in defense response to bacterium
GGT1	Gamma-glutamyltranspeptidase	Involved in regulation of inflammatory response
PCK1	Soluble phosphoenolpyruvate carboxykinase 1	Involved in cellular response to glucose stimulus
HSP70b	Heat-shock protein 70b	Enables ATP binding
RCI3	Peroxidase 3	Part of plant-type cell wall, involved in response to oxidative stress

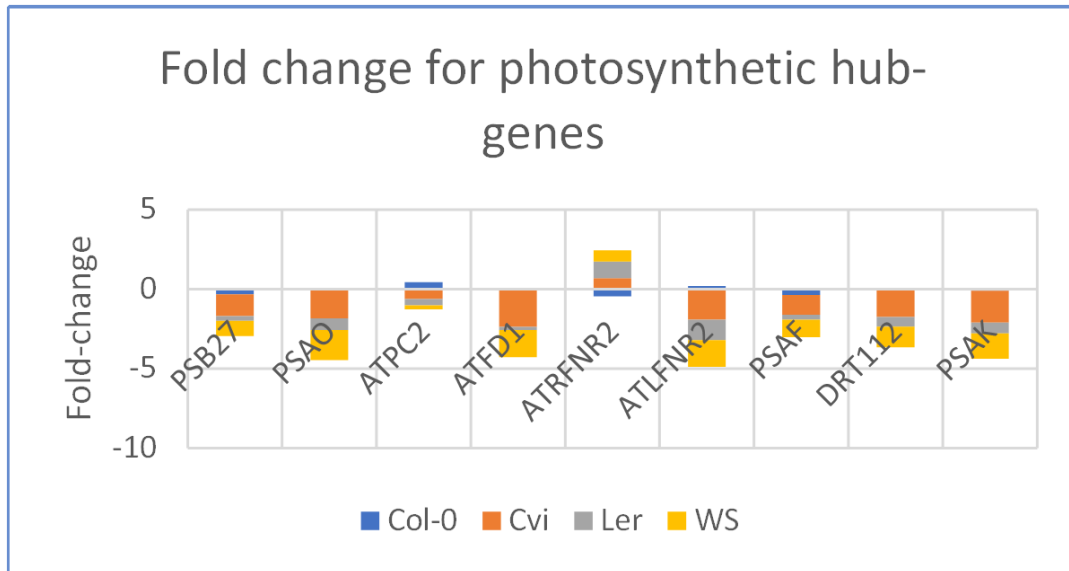
3.2. Identification of Stress Response Genes in Grn of Spaceflight Microgravity and Ground Control

We performed Gene ontology on the genes to determine the stress response genes in photosynthesis and carbon fixation GRN (refer to Table 1). The photosynthetic stress response genes PETC, ATPD, ATLFNR1, and ATLFNR2 are seen in all ecotypes. These stress response mechanisms include response to bacterium and other organisms, biotic stimulus, and defense for all the ecotypes. The stress response mechanisms in the carbon fixation biological process include response to temperature stimulus, stress, cold, abiotic stimulus, and biotic stimulus. The genes that correspond to these stress response mechanisms are PGK1, GAPB, PRK, GAPC, MDH, ATCTIMC, PCK1, SBPASE, ASP1, and GGT1. All of these genes are not observed in all the GRN. The reason is that the Lasso regression method eliminates some of the minimally correlated genes.

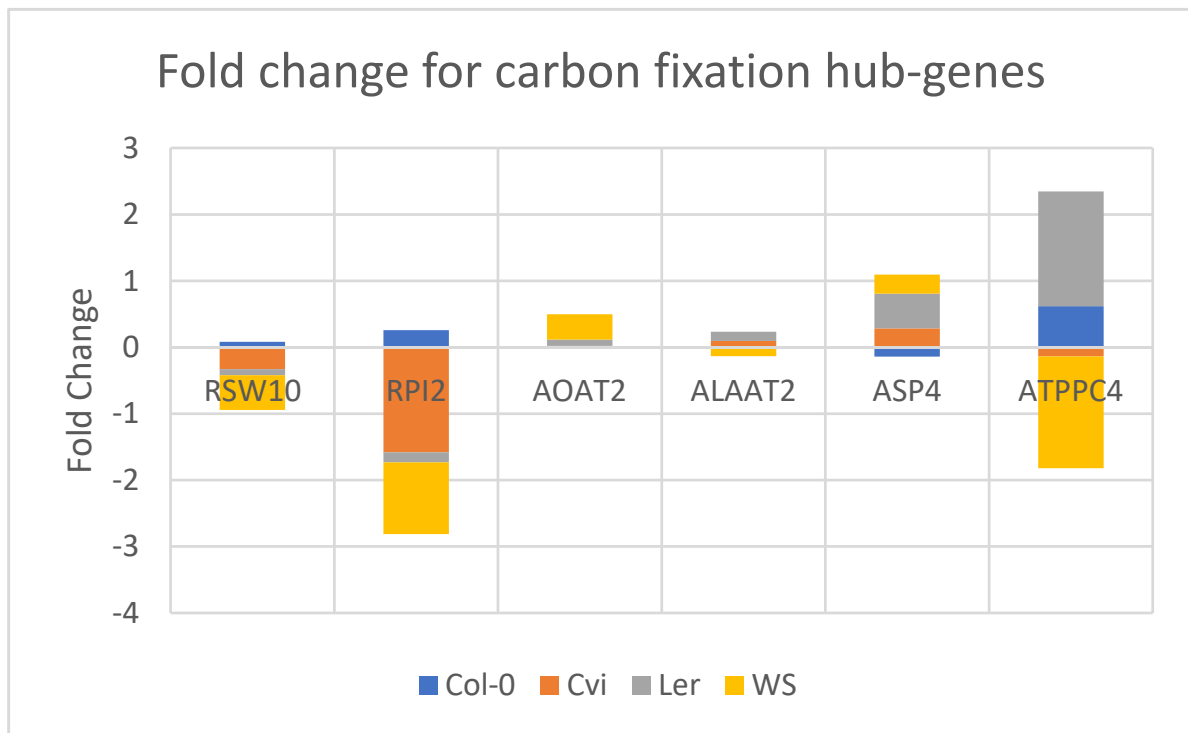
3.3. Photosynthesis Genes Are Downregulated in Spaceflight Microgravity

We have analyzed common hub-genes and stress response genes in the networks to compare their regulations in spaceflight microgravity and ground control. The fold-change analysis of common photosynthetic hub-genes is shown in Figure 3A. This analysis depicts that most of the photosynthetic hub-genes are downregulated in spaceflight microgravity. The fold-change analysis of carbon fixation hub-genes (see Figure 3B) reveals that the genes

are both upregulated and downregulated in the ecotypes. The fold-change analysis of photosynthetic stress response genes (see Figure 3C) reveals that the genes are downregulated in spaceflight microgravity except for the gene ATLFNR2, which is upregulated in the Col-0 ecotype. The fold-change analysis of stress response genes in carbon fixation genes (see Figure 3D) reveals that GAPC, GGT1, and MDH are upregulated in most of the ecotypes in spaceflight microgravity.

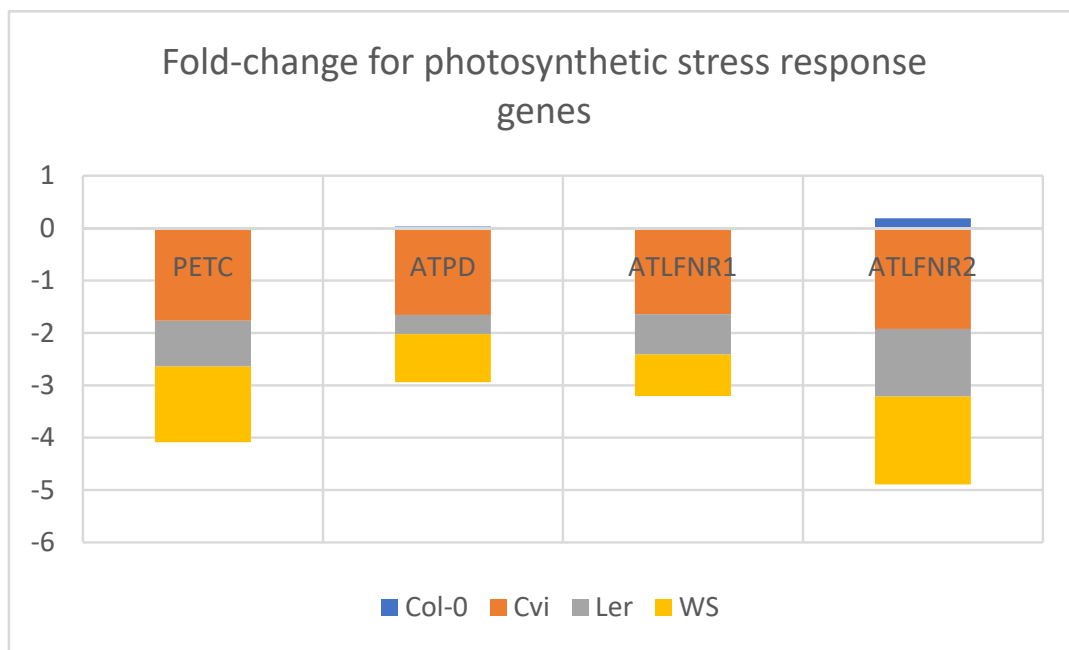


(A)

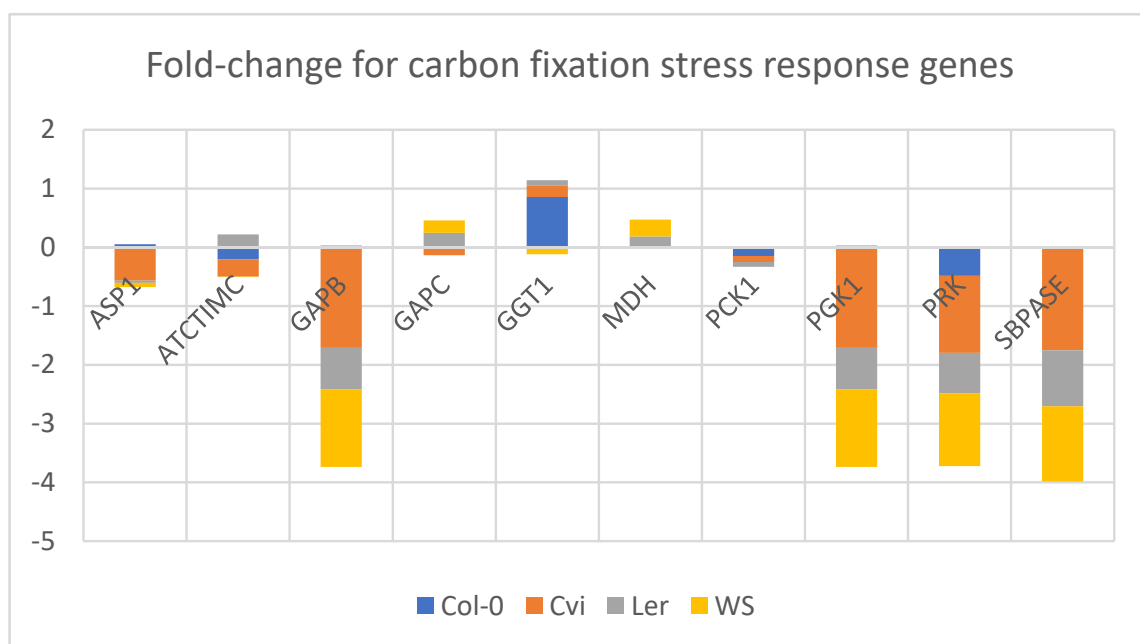


(B)

Figure 3. Cont.



(C)



(D)

Figure 3. Fold-change analyses of the genes in the different ecotypes of *Arabidopsis* in spaceflight microgravity when compared to ground control. (A) Fold-change for photosynthetic hub-genes for spaceflight microgravity vs. ground control. (B) Fold-change for carbon fixation hub-genes for spaceflight microgravity vs. ground control. (C) Fold-change for photosynthetic stress-response genes for spaceflight microgravity vs. ground control. (D) Fold-change for carbon fixation stress-response genes for spaceflight microgravity vs. ground control.

3.4. *Cvi* Ecotype Has the Same Outdegree Distributions in Spaceflight Microgravity and Ground Control

The transcriptional regulations (outdegree distributions) of common hub-genes in photosynthesis and carbon fixation biological processes are different across each of the GRN

in spaceflight microgravity and ground control. The outdegree distributions of common hub-genes are displayed in Figure 4A. It is observed that gene PSB27, which plays a vital role in the metabolic activity of the photosynthesis process, has maximum outdegree in all the GRN. The outdegree distributions of the photosynthetic and carbon fixation stress response genes are displayed in Figure 4B. The figure reveals that the photosynthetic stress response gene PETC has maximum outdegree in all the GRN. As mentioned earlier (see Section 2.2), some of the carbon fixation stress response genes are not observed in all the GRN. Hence, one cannot find the outdegree distribution. The other reason is that some of the genes have no outdegree in the GRN. They have an in-degree, which means they are influenced by other transcriptional factors (hub-genes) as a response to stress for the survival of the plant. The Cvi ecotype has the same outdegree in the GRN of spaceflight microgravity and ground control.

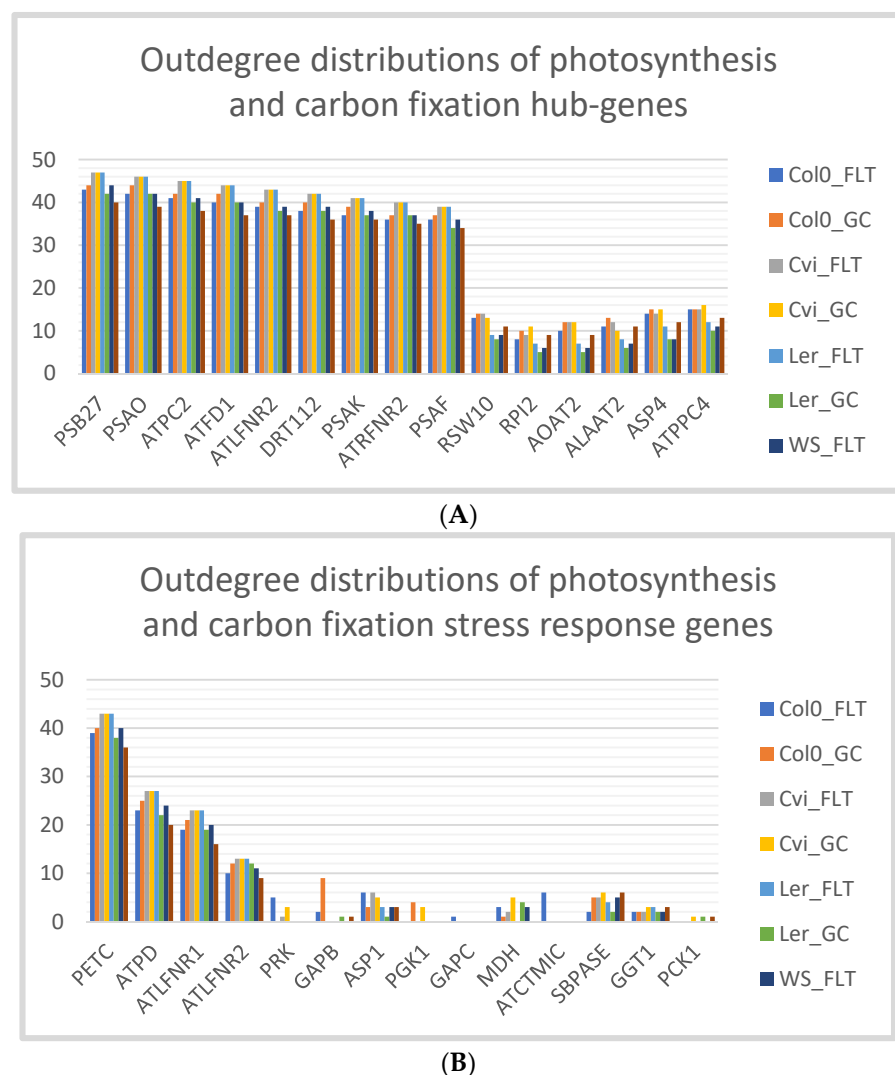


Figure 4. Outdegree distributions of photosynthesis and carbon fixation genes in all ecotypes in spaceflight microgravity and ground control. (A) Outdegree distributions of photosynthesis and carbon fixation hub-genes. (B) Outdegree distributions of photosynthesis and carbon fixation stress response genes.

3.5. Stress Response Genes of Col-0 Ecotype Have Low Shortest Path Lengths in Spaceflight Microgravity

The shortest path lengths for hub-genes and stress response genes are calculated and compared. Figure 5A displays the comparison of average shortest path lengths in different

ecotypes for the hub-genes. AOAT2 and ALAAT2 have the highest shortest path length in Ler ecotype in ground control. The lower shortest path lengths are observed in the genes in Col-0 ecotype in spaceflight microgravity when compared to other ecotypes in spaceflight microgravity. The hub-genes in all ecotypes in ground control have lesser shortest path lengths compared to spaceflight microgravity except for a few genes in the Ler ecotype. Most of the stress response genes in the Col-0 ecotype have very low shortest path length in spaceflight microgravity (see Figure 5B). GAPB gene has the highest shortest path length in Col-0 ecotype in spaceflight microgravity. Most of the stress response genes have an average shortest path length of one in all the ecotypes in spaceflight microgravity and ground control.

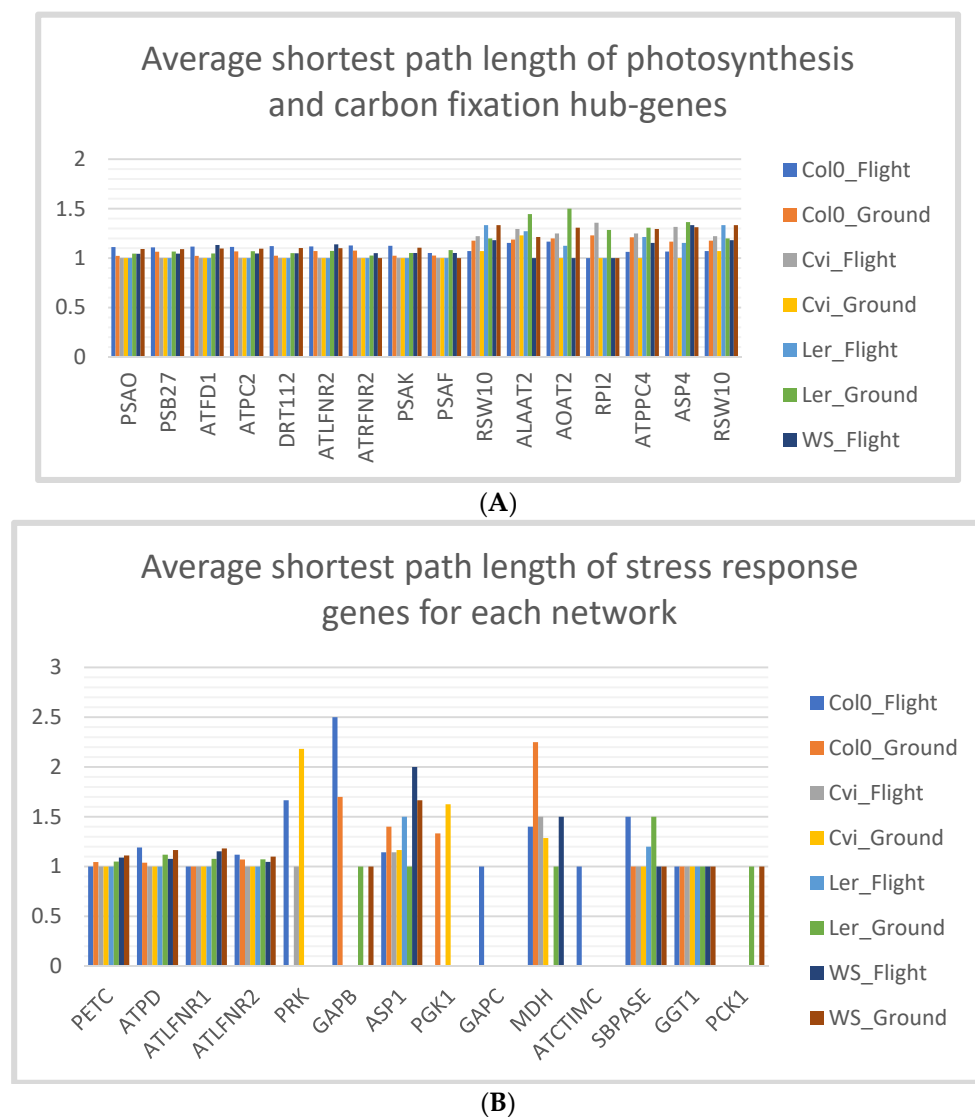


Figure 5. Average shortest path lengths of photosynthesis and carbon fixation genes in all ecotypes in spaceflight microgravity and ground control. **(A)** Average shortest path lengths of photosynthesis and carbon fixation hub-genes. **(B)** Average shortest path lengths of photosynthesis and carbon fixation stress response genes.

3.6. Photosynthesis Hub-Genes Have Low Betweenness Centrality

The comparison of betweenness centrality of hub-genes and stress response genes of the ecotypes in spaceflight microgravity and ground control are shown in Figure 6. Most of the photosynthesis hub-genes in all the GRN have betweenness centrality close to 0 (see Figure 6A). The carbon fixation hub-genes have higher betweenness centrality in ground

control compared to spaceflight microgravity. The highest betweenness centrality in stress response genes is observed in WS ecotype in ground control (see Figure 6B). ASP1 gene in the GRN of Ler ecotype in spaceflight microgravity has higher betweenness centrality when compared to other genes and ecotypes in spaceflight microgravity.

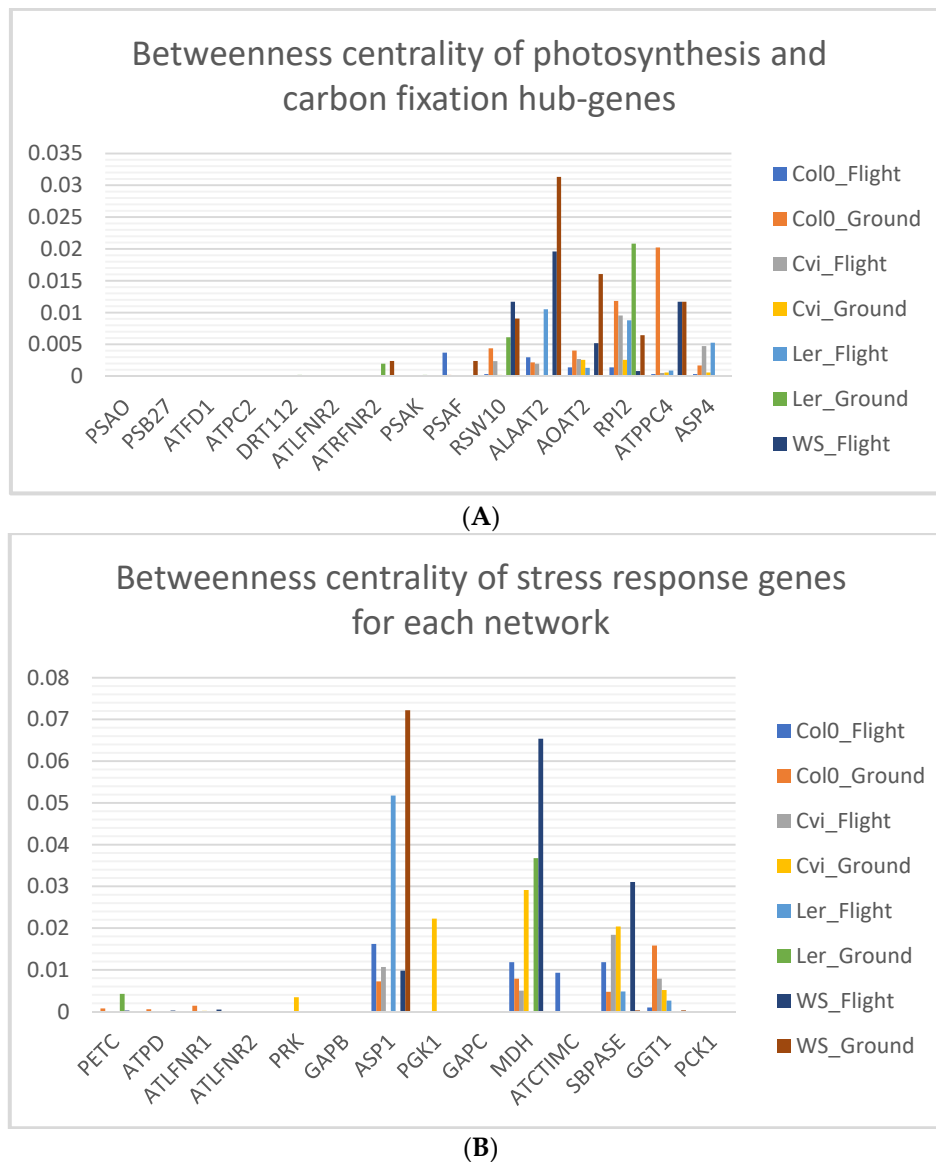
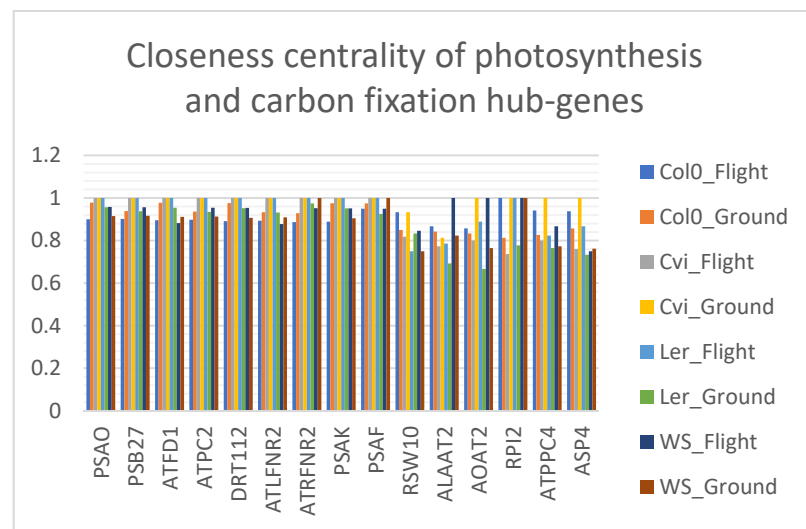


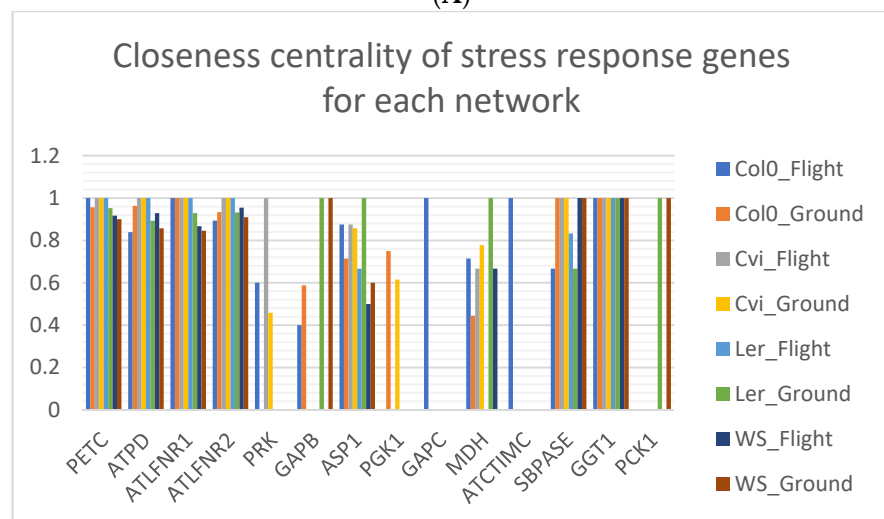
Figure 6. Betweenness centrality of photosynthesis and carbon fixation genes in all ecotypes in spaceflight microgravity and ground control. **(A)** Betweenness centrality of photosynthesis and carbon fixation hub-genes. **(B)** Betweenness centrality of photosynthesis and carbon fixation stress response genes.

3.7. Closeness Centrality Is Lowered in Spaceflight Microgravity

The comparison of closeness centrality of hub-genes and stress response genes of the ecotypes in spaceflight microgravity and ground control are shown in Figure 7. The hub-genes in ground control have a higher closeness centrality when compared to spaceflight microgravity (see Figure 7A). The same behavior of stress response genes has been observed (see Figure 7B) in spaceflight microgravity. The carbon fixation stress response genes have a very low closeness centrality in spaceflight microgravity.



(A)



(B)

Figure 7. Closeness centrality of photosynthesis and carbon fixation genes in all ecotypes in spaceflight microgravity and ground control. (A) Closeness centrality of photosynthesis and carbon fixation hub-genes. (B) Closeness centrality of photosynthesis and carbon fixation stress response genes.

3.8. Col-0 Ecotype Hub-Genes Have High Eccentricity in Spaceflight Microgravity Compared to Ground Control

The comparison of the eccentricity of hub-genes and stress response genes of the ecotypes in spaceflight microgravity and ground control are shown in Figure 8. The photosynthesis hub-genes of the Cvi ecotype in spaceflight microgravity and ground control have a very low eccentricity compared to other ecotypes (see Figure 8A). The Ler ecotype in spaceflight microgravity has the same behavior of photosynthesis hub-genes as that of the Cvi ecotype. The stress response genes in the GRN of Col-0 ecotype in spaceflight microgravity have low eccentricity when compared to another ecotype except for the gene GAPB (see Figure 8B). The gene GAPB has the highest eccentricity in the Col-0 ecotype in spaceflight microgravity. The eccentricity of 1 is observed in most of the stress response genes in all GRN. None of the genes have an eccentricity of 0.

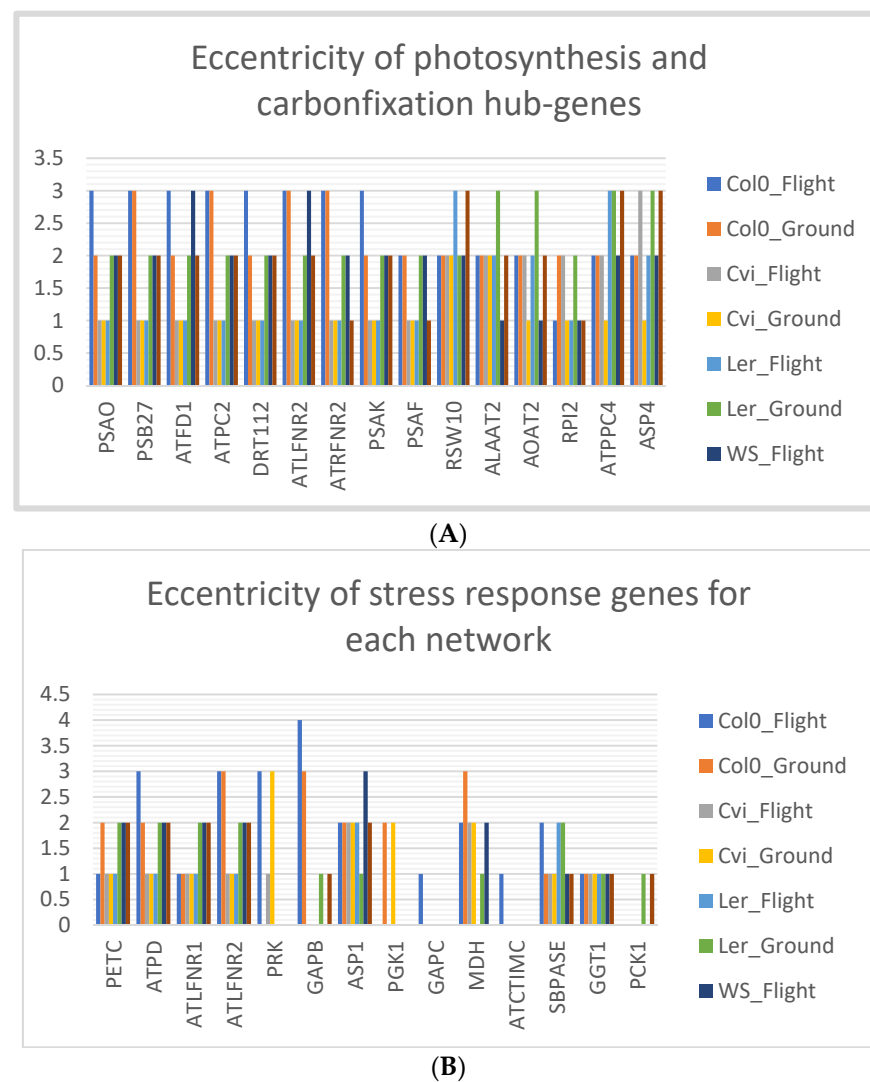


Figure 8. Eccentricity of photosynthesis and carbon fixation genes in all ecotypes in spaceflight microgravity and ground control. **(A)** The eccentricity of photosynthesis and carbon fixation hub-genes. **(B)** The eccentricity of photosynthesis and carbon fixation stress response genes.

3.9. Interactions of Oxidative Stress Response Genes with Photosynthesis Hub-Genes

We have extracted the sub-network of oxidative stress response genes that interact with the photosynthesis genes from the whole network. The heat-shock protein HSP70b and the novel cold-inducible gene RCI3 are the two genes that interact with the photosynthetic genes in all the ecotypes in spaceflight microgravity and ground control. The fold-change analysis revealed that HSP70b is upregulated in all the ecotypes in spaceflight microgravity, and RCI3 is downregulated in all ecotypes in spaceflight microgravity. The interactions of HSP70b and RCI3 genes with photosynthetic genes are shown in Figure 9. The genes HSP70b and RCI3 are interacting only with the genes ATFD1, ATLFNR2, DRT112, ATRFNR2, and PSAO that play a vital role in the oxidation-reduction process (refer to Table 1). HSP70b and RCI3 interact with the gene PSB27 that is involved in light-harvesting and reaction mechanisms to light in the photosynthetic process. These genes also interact with ATPC2 that regulates ATPase activity and the genes PSAF and PSAK that play a significant role in PSI (refer to Table 1). The interactions of these genes are common across all the ecotypes in both environments. The comparison of the edge-list of networks disclosed that genes are differentially regulated between spaceflight and ground environments in all the ecotypes. The Col-0 ecotype has a large number of regulations in the GRN in spaceflight microgravity when compared with the GRN of ground control. The other three ecotypes

have a greater number of inhibitions in the GRN of spaceflight microgravity compared to the GRN of ground control.

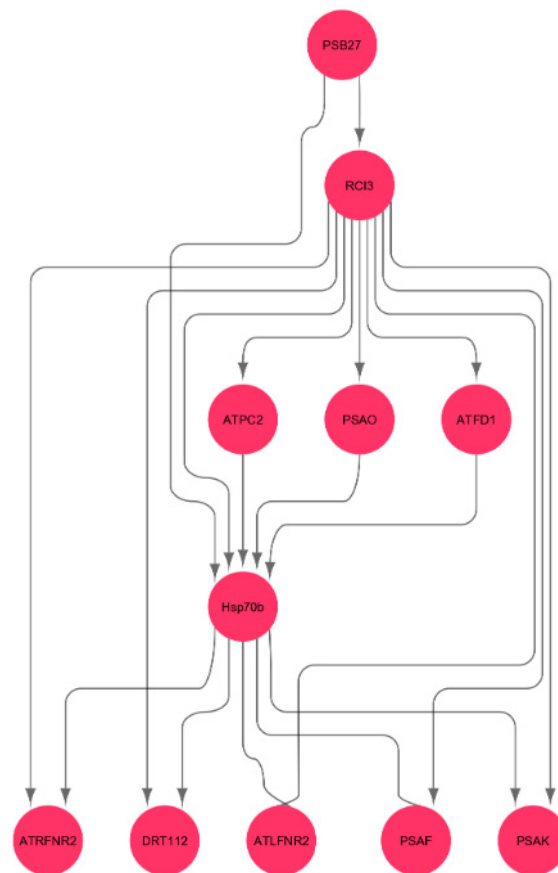


Figure 9. Sub-network of interactions of HSP70b and RCI3 genes with the photosynthesis genes.

4. Discussion

The topological properties of the GRN of four *Arabidopsis* ecotypes in spaceflight microgravity and ground control are discussed here. GRNs are useful for linking TFs (hub-genes) to their target genes, thereby representing transcriptional gene regulations as a graph (network). The individual behavior of the genes (nodes) has an impact on the small-world or scale-free phenomena of the network [17]. The shortest path length, centrality (betweenness and closeness), and eccentricity of photosynthesis hub-genes and stress response genes are the topological measures calculated on the individual GRN.

4.1. Low Shortest Path Length Indicates Small-World-Ness of the Network

Figure 5 shows low as well as high values of the average shortest path length. The low values for average shortest path length are an indication of a small-world network, and the shortest path length of each gene is essential to achieve the small world-ness of the GRN [29]. The analysis of shortest path lengths of both stress response genes and hub-genes reveals that the genes have a higher value for shortest path length in spaceflight microgravity compared to ground control, which is an indication that the photosynthetic GRN might lose the small-world-ness in spaceflight microgravity.

4.2. High Network Centrality Indicates the Importance of Genes on the Whole Network

Network centrality is an index that shows which node in the network has a critical position in the whole network in connecting with other nodes [50]. In this paper, we have analyzed betweenness centrality and closeness centrality of the photosynthetic hub-genes and stress response genes of all ecotypes in spaceflight microgravity and ground control.

The higher the value of betweenness centrality, the gene occurs more often between the shortest paths of the other genes [50]. Some hub-genes have a betweenness centrality of 0 because they have an outdegree but no indegree (see Figure 6). Hence, they do not act as a bridge between the other genes indicating that they act only as transcription factors but not target genes.

The closeness centrality value ranges between 0 and 1. As discussed in [50], higher closeness centrality (close to 1) is an indication that the gene is fully connected in the network. The hub-genes of the Cvi ecotype in both the environments have a closeness centrality of 1, indicating that they are fully connected in the network (see Figure 7). The GRN of LER ecotype in spaceflight microgravity has all the hub-genes fully connected in the network. The Col-0 ecotype in spaceflight microgravity has no hub-gene with a closeness centrality of 1. This is because the stress response genes regulate most of the genes to adapt to the spaceflight environment.

4.3. High Eccentricity Indicates Higher Connections in the Network

As discussed in [32], the higher the eccentricity of the node, the greater is its connections with other nodes in the network. The center of the network tends to have minimum eccentricity for all nodes. Hence, the node with higher eccentricity has an indirect impact on the whole network. The hub-genes in spaceflight microgravity have higher eccentricity when compared to ground control because they have a greater outdegree in spaceflight microgravity when compared to the outdegree of the same genes in ground control. All the stress response genes in spaceflight microgravity have an eccentricity of 1. The GAPB in the Col-0 ecotype in spaceflight microgravity has the highest eccentricity of 4.

4.4. Heat Shock Gene Regulates Photosynthesis Genes in Spaceflight Microgravity

The analysis of the individual behavior of the genes in the GRN is essential for understanding the collective impact of the genes on the GRN. We have noticed that the stress response genes of the ecotypes are not present in the GRN of both the spaceflight and ground environments. This is because the Lasso regression eliminates minimal correlation (close to 0) of the genes [51]. The minimal correlation of the stress response genes is because of two reasons: there might not be much needed for the genes to express in response to spaceflight microgravity, or the genes exhibit dysfunctionality in the spaceflight microgravity environment [4]. The dysfunctionality of the genes in photosynthesis might be because of the oxidative stress that occurs due to the production of harmful oxygen species when the plants are exposed to high radiation [13].

The heat-shock protein HSP70b is upregulated in spaceflight microgravity and interacts with the photosynthetic genes PSB27, ATPC2, PSAO, ATFD1, ATEFNR2, DRT112, ATLFNR2, PSAF, and PSAK. The previous studies on HSP70b show that the gene is responsible for repairing the genes that belong to PSII (refer to Table 1) to reduce oxidative stress [52]. As discussed in [53], the downregulation of HSP70b results in the photo-sensitivity of the plant, whereas, upregulation of the gene has a protective effect on the plant. The upregulation of HSP70b in spaceflight microgravity is an indication that the gene provides a defense mechanism with the help of the genes involved in the oxidation-reduction process for the survival of *Arabidopsis* in spaceflight microgravity.

4.5. Spaceflight Environment Leads to Dehydration-And-Salt-Stress Sensitive Ecotypes

A detailed description of RCI3 is mentioned in [54], where the upregulation of the gene showed an increased tolerance of the plant in dehydration and salt stress. The downregulation of the gene resulted in a dehydration-and-salt sensitive plant. RCI3 is downregulated in all the ecotypes of *Arabidopsis* in spaceflight microgravity and interacts with the photosynthetic hub-genes. There is a chance of dehydration stress in the spaceflight environment as the availability of water is less compared to the ground environment. The downregulation of RCI3 is an indication that the plants are sensitive to dehydration in spaceflight microgravity.

4.7. Comparison of the Stress Response Genes in Photosynthesis and Carbon Fixation Processes of *Arabidopsis thaliana* under Different Stress Conditions

Spaceflight conditions introduce other stressors such as low atmospheric pressure, low oxygen, and higher doses of radiation to which plants have to adapt. It is interesting to compare how the stress response genes involved in photosynthesis and the carbon fixation process are affected by other environmental stressors in spaceflight. For performing this comparison, we have selected two GeneLab datasets, GLDS-46 and GLDS-136. The dataset GLDS-46 contains the transcriptomic responses of *Arabidopsis* when exposed to radiation. Two different kinds of *Arabidopsis* are considered for the experiment: wild-type and mutants defective in DSB-sensing protein kinase ATM. These two types of plants are exposed to different ionizing radiation treatment types, HZE, and gamma photons. The responses of the plants are compared with the plants grown under control conditions. The complete description of the mission can be found in [56]. The transcriptomic responses of the WS ecotype of *Arabidopsis* under hypobaric and hypoxia conditions are available in the GLDS-136 dataset. A fold change analysis of the photosynthetic stress response genes of WT and ATM mutant plants after exposure to gamma and HZE radiation for 24 h is conducted and presented in Figure 11. The complete experimental setup can be found in [57].

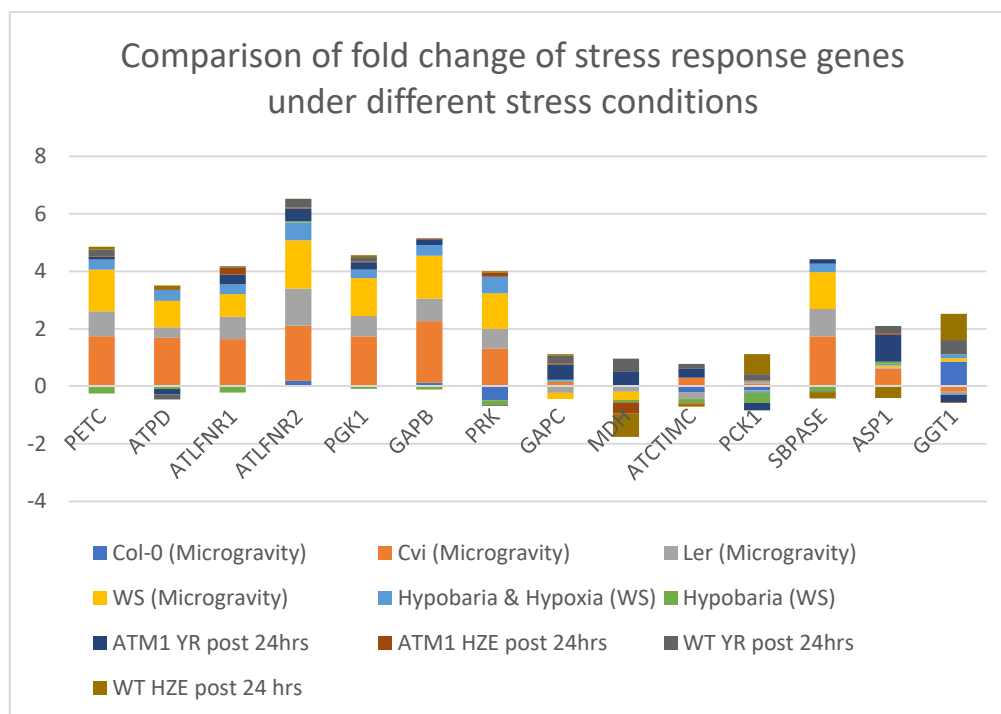


Figure 11. Comparison of the stress response genes in photosynthesis and carbon fixation processes of *Arabidopsis* under different stress conditions.

We have made the fold change analysis of photosynthetic stress response genes under hypobaric and hypobaric + hypoxia conditions compared to normal atmospheric pressure and oxygen conditions. Most of the genes are upregulated under all stress conditions. Responses to *Arabidopsis* under salinity stress and other stresses are also studied in [58,59].

5. Conclusions

Graph-theoretic network analysis performed on the transcriptional gene expression datasets of four different ecotypes of *Arabidopsis* has brought to light significant gene regulations in spaceflight that are similar or different compared to ground control. While earlier investigations have relied on fold change analysis alone for identifying individual key gene players in spaceflight microgravity, here, we have applied network analysis approaches

for identifying hub genes as well as the interactor genes and processes associated with these hub genes. The topological analysis of the GRN reveals the individual behavior of the genes in spaceflight microgravity and how it impacts the overall photosynthetic functioning of the plant in the altered spaceflight environment. Photosynthetic plants like *Arabidopsis* obtain defense mechanisms by upregulating specific genes for the survival of plants in adverse conditions like spaceflight microgravity, thereby ensuring that the normal photosynthetic process takes place in the altered environments. It is also seen that these genes are upregulated under typical spaceflight environmental conditions such as low atmospheric pressure, low oxygen, and higher doses of radiation.

Supplementary Materials: The GeneLab datasets are online at www.genelab.nasa.gov. The scripts are available at <https://github.com/gharshini/Network-Analysis-of-Local-Gene-Regulators-in-Arabidopsis-Thaliana-Under-Spaceflight-Stress->.

Author Contributions: Conceptualization, H.G., V.M., methodology, H.G., V.M., J.O., H.J., C.A.; software, H.G., J.O., C.A.; validation, H.G., J.O., C.A.; formal analysis, H.G., J.O., V.M., H.J.; investigation, V.M., H.J.; data curation, H.G., J.O., C.A.; writing—original draft preparation, V.M., H.G., H.J.; writing—review and editing, V.M. and H.J.; visualization, H.G., J.O., V.M.; supervision, V.M. and H.J.; project administration, V.M.; funding acquisition, V.M. and H.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by NASA EPSCoR, grant number 80NSSC19M0167. The APC is funded by 80NSSC19M0167. Opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NASA.

Institutional Review Board Statement: Not Applicable.

Informed Consent Statement: Not Applicable.

Data Availability Statement: <https://www.genelab.nasa.gov/>.

Acknowledgments: Carlos A. Agrinoni's work is supported by the NASA Training Grant No. NNX15AI11H. Opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NASA.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Koornneef, M.; Meinke, D. The development of *Arabidopsis* as a model plant. *Plant J.* **2010**, *61*, 909–921. [[CrossRef](#)] [[PubMed](#)]
2. Initiative, T.A.G. Analysis of the genome sequence of *Arabidopsis thaliana*. *Nature* **2000**, *408*, 796–815. [[CrossRef](#)] [[PubMed](#)]
3. Oakley, C.G.; Savage, L.; Lotz, S.; Larson, G.R.; Thomashow, M.F.; Kramer, D.M.; Schemske, D.W. Genetic basis of photosynthetic responses to cold in two locally adapted populations of *Arabidopsis thaliana*. *J. Exp. Bot.* **2018**, *69*, 699–709. [[CrossRef](#)] [[PubMed](#)]
4. Paul, A.-L.; Sng, N.J.; Zupanska, A.K.; Krishnamurthy, A.; Schultz, E.R.; Ferl, R.J. Genetic dissection of the *Arabidopsis* spaceflight transcriptome: Are some responses dispensable for the physiological adaptation of plants to spaceflight? *PLoS ONE* **2017**, *12*, e0180186. [[CrossRef](#)] [[PubMed](#)]
5. Bouchabke, O.; Chang, F.; Simon, M.; Voisin, R.; Pelletier, G.; Durand-Tardif, M. Natural Variation in *Arabidopsis thaliana* as a Tool for Highlighting Differential Drought Responses. *PLoS ONE* **2008**, *3*, e1705. [[CrossRef](#)] [[PubMed](#)]
6. Barah, P.; Jayavelu, N.D.; Mundy, J.; Bones, A.M. Genome scale transcriptional response diversity among ten ecotypes of *Arabidopsis thaliana* during heat stress. *Front. Plant Sci.* **2013**, *4*. [[CrossRef](#)]
7. Wang, Y.; Yang, L.; Zheng, Z.; Grumet, R.; Loescher, W.; Zhu, J.-K.; Yang, P.; Hu, Y.; Chan, Z. Transcriptomic and Physiological Variations of Three *Arabidopsis* Ecotypes in Response to Salt Stress. *PLoS ONE* **2013**, *8*, e69036. [[CrossRef](#)]
8. Prasad, C.M.; Sonnewald, U. In silico selection of *Arabidopsis thaliana* ecotypes with enhanced stress tolerance. *Plant Signal. Behav.* **2013**, *8*, e26364. [[CrossRef](#)]
9. Li, P.; Mane, S.P.; Sioson, A.A.; Vasquez-Robinet, C.; Heath, L.S.; Bohnert, H.J.; Grene, R. Effects of chronic ozone exposure on gene expression in *Arabidopsis thaliana* ecotypes and in *Thellungiella halophila*. *Plant Cell Environ.* **2006**, *29*, 854–868. [[CrossRef](#)]
10. Kakouri, A.C.; Christodoulou, C.C.; Zachariou, M.; Oulas, A.; Spyrou, G.M.; Demetriou, C.A.; Votsi, C.; Zamba-Papanicolaou, E.; Christodoulou, K.; Spyrou, G.M. Revealing Clusters of Connected Pathways Through Multisource Data Integration in Huntington's Disease and Spastic Ataxia. *IEEE J. Biomed. Health Inform.* **2018**, *23*, 26–37. [[CrossRef](#)]
11. Choi, W.-G.; Barker, R.J.; Kim, S.-H.; Swanson, S.J.; Gilroy, S. Variation in the transcriptome of different ecotypes of *Arabidopsis thaliana* reveals signatures of oxidative stress in plant responses to spaceflight. *Am. J. Bot.* **2019**, *106*, 123–136. [[CrossRef](#)]
12. Leister, D.; Schneider, A. From Genes to Photosynthesis in *Arabidopsis thaliana*. *Adv. Virus Res.* **2003**, *228*, 31–83.

13. Yokthongwattana, K.; Chrost, B.; Behrman, S.; Casper-Lindley, C.; Melis, A. Photosystem II Damage and Repair Cycle in the Green Alga *Dunaliella salina*: Involvement of a Chloroplast-Localized HSP70. *Plant Cell Physiol.* **2001**, *42*, 1389–1397. [[CrossRef](#)] [[PubMed](#)]
14. Emmert-Streib, F.; Dehmer, M.; Haibe-Kains, B. Gene regulatory networks and their applications: Understanding biological and medical problems in terms of networks. *Front. Cell Dev. Biol.* **2014**, *2*, 38. [[CrossRef](#)] [[PubMed](#)]
15. Subramanian, A.; Tamayo, P.; Mootha, V.K.; Mukherjee, S.; Ebert, B.L.; Gillette, M.A.; Paulovich, A.; Pomeroy, S.L.; Golub, T.R.; Lander, E.S.; et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15545–15550. [[CrossRef](#)]
16. Chen, L.; Chu, C.; Lu, J.; Kong, X.; Huang, T.; Cai, Y.-D. Gene Ontology and KEGG Pathway Enrichment Analysis of a Drug Target-Based Classification System. *PLoS ONE* **2015**, *10*, e0126492. [[CrossRef](#)]
17. Ma, X.; Gao, L. Biological network analysis: Insights into structure and functions. *Briefings Funct. Genom.* **2012**, *11*, 434–442. [[CrossRef](#)]
18. Janwa, H.; Massey, S.E.; Velez, J.; Mishra, B. On the Origin of Biomolecular Networks. *Front. Genet.* **2019**, *10*, 240. [[CrossRef](#)]
19. Kanehisa, M.; Furumichi, M.; Tanabe, M.; Sato, Y.; Morishima, K. KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **2017**, *45*, D353–D361. [[CrossRef](#)]
20. Lenz, M.; Müller, F.-J.; Zenke, M.; Schuppert, A. Principal components analysis and the reported low intrinsic dimensionality of gene expression microarray data. *Sci. Rep.* **2016**, *6*, 25696. [[CrossRef](#)]
21. Raychaudhuri, S.; Stuart, J.M.; Altman, R.B. Principal components analysis to summarize microarray experiments: Application to sporulation time series. *Bioinformatics* **2001**, *1999*, 455–466.
22. Chai, L.E.; Loh, S.K.; Low, S.T.; Mohamad, M.S.; Deris, S.; Zakaria, Z. A review on the computational approaches for gene regulatory network construction. *Comput. Biol. Med.* **2014**, *48*, 55–65. [[CrossRef](#)] [[PubMed](#)]
23. Hempel, S.; Koseska, A.; Nikoloski, Z.; Kurths, J. Unraveling gene regulatory networks from time-resolved gene expression data—A measures comparison study. *BMC Bioinform.* **2011**, *12*, 292. [[CrossRef](#)] [[PubMed](#)]
24. Omranian, N.; Eloundou-Mbebi, J.M.O.; Mueller-Roeber, B.; Nikoloski, Z. Gene regulatory network inference using fused LASSO on multiple data sets. *Sci. Rep.* **2016**, *6*, 20533. [[CrossRef](#)]
25. Zhang, R.; Ren, Z.; Chen, W. SILGGM: An extensive R package for efficient statistical inference in large-scale gene networks. *PLoS Comput. Biol.* **2018**, *14*, e1006369. [[CrossRef](#)]
26. Sartor, M.A.; Leikauf, G.D.; Medvedovic, M. LRpath: A logistic regression approach for identifying enriched biological groups in gene expression data. *Bioinformatics* **2008**, *25*, 211–217. [[CrossRef](#)]
27. Quackenbush, J. Microarray data normalization and transformation. *Nat. Genet.* **2002**, *32*, 496–501. [[CrossRef](#)]
28. Albert, R.; Barabási, A.-L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **2002**, *74*, 47–97. [[CrossRef](#)]
29. Mao, G.; Zhang, N. Analysis of Average Shortest-Path Length of Scale-Free Network. *J. Appl. Math.* **2013**, *2013*, 1–5. [[CrossRef](#)]
30. Freeman, L.C. A Set of Measures of Centrality Based on Betweenness. *Sociometry* **1977**, *40*. [[CrossRef](#)]
31. Freeman, L.C. Centrality in social networks conceptual clarification. *Soc. Netw.* **1978**, *1*, 215–239. [[CrossRef](#)]
32. Krnc, M.; Sereni, J.-S.; Skrekovski, R.; Yilma, Z. Eccentricity of Networks with Structural Constraints. *Discuss. Math. Graph Theory* **2020**, *40*, 1141–1162. [[CrossRef](#)]
33. Watts, D.J.; Strogatz, S.H. Collective dynamics of “small world” networks. *Nature* **1998**, *393*, 440–442. [[CrossRef](#)] [[PubMed](#)]
34. Loscalzo, J.; Barabási, A.-L. *Network Science*, 1st ed.; Cambridge University Press: Cambridge, UK, 2016.
35. Newman, M.E. Assortative mixing in networks. *Phys. Rev. Lett.* **2002**, *89*, 208701. [[CrossRef](#)]
36. Kleinberg, J.M. Authoritative sources in a hyperlinked environment. *J. ACM* **1999**, *46*, 604–632. [[CrossRef](#)]
37. Lei, S.; She, Y.; Zeng, J.; Chen, R.; Zhou, S.; Shi, H. Expression patterns of regulatory lncRNAs and miRNAs in muscular atrophy models induced by starvation in vitro and in vivo. *Mol. Med. Rep.* **2019**, *20*, 4175–4185. [[CrossRef](#)]
38. Easley, D.; Kleinberg, J. *Networks, Crowds, and Markets*; Cambridge University Press: Cambridge, UK, 2010.
39. Golub, G.H.; Van Loan, C.F. *Matrix Computations (Johns Hopkins Studies in Mathematical Sciences)*, 3rd ed.; The Johns Hopkins University Press: Baltimore, MD, USA, 1996.
40. NetworkX. Available online: <https://networkx.org> (accessed on 30 August 2019).
41. Miller, J.C.; Rae, G.; Schaefer, F.; Ward, L.A.; LoFaro, T.; Farahat, A. Modifications of Kleinberg’s HITS algorithm using matrix exponentiation and web log records. In Proceedings of the 24th annual international ACM SIGIR conference on Research and Development in Information Retrieval, New Orleans, LA, USA, 9–12 September 2001; pp. 444–445.
42. Langville, A.N.; Meyer, C.D. A survey of eigenvector methods for web information retrieval. *SIAM Rev.* **2005**, *47*, 135–161. [[CrossRef](#)]
43. Emms, D.M.; Kelly, S. OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **2015**, *16*, 157. [[CrossRef](#)]
44. Bian, T.; Hu, J.; Deng, Y. Identifying influential nodes in complex networks based on AHP. *Phys. A Stat. Mech. Appl.* **2017**, *479*, 422–436. [[CrossRef](#)]
45. Urbinati, A.; Galimberti, E.; Ruffo, G. Hubs and authorities of the scientific migration network. *arXiv* **2019**, arXiv:1907.07175.
46. Eldén, L. *Matrix Methods in Data Mining and Pattern Recognition*; SIAM: Philadelphia, PA, USA, 2019; ISBN 978-0-89871-626-9.
47. Cowen, L.; Ideker, T.; Raphael, B.J.; Sharan, R. Network propagation: A universal amplifier of genetic associations. *Nat. Rev. Genet.* **2017**, *18*, 551. [[CrossRef](#)] [[PubMed](#)]

48. Opsahl, T.; Agneessens, F.; Skvoretz, J. Node centrality in weighted networks: Generalizing degree and shortest paths. *Soc. Net.* **2010**, *32*, 245–251. [[CrossRef](#)]
49. Gazestani, V.H.; Lewis, N.E. From Genotype to Phenotype: Augmenting Deep Learning with Networks and Systems Biology. *Curr. Opin. Syst. Biol.* **2019**. [[CrossRef](#)] [[PubMed](#)]
50. Zhang, J.; Luo, Y. Degree Centrality, Betweenness Centrality, and Closeness Centrality in Social Network. *Adv. Intelligent Syst. Res.* **2017**, *132*, 300–303. [[CrossRef](#)]
51. Wang, S.; Nan, B.; Rosset, S.; Zhu, J. Random lasso. *Bone* **2008**, *23*, 1–7. [[CrossRef](#)] [[PubMed](#)]
52. Sung, D.Y.; Vierling, E.; Guy, C.L. Comprehensive Expression Profile Analysis of the Arabidopsis Hsp70 Gene Family. *Plant Physiol.* **2001**, *126*, 789–800. [[CrossRef](#)] [[PubMed](#)]
53. Schroda, M.; Vallon, O.; Wollman, F.-A.; Beck, C.F. A Chloroplast-Targeted Heat Shock Protein 70 (HSP70) Contributes to the Photoprotection and Repair of Photosystem II during and after Photoinhibition. *Plant Cell* **1999**, *11*, 1165–1178. [[CrossRef](#)] [[PubMed](#)]
54. Llorente, F.; López-Cobollo, R.M.; Catalá, R.; Martínez-Zapater, J.M.; Salinas, J. A novel cold-inducible gene from Arabidopsis, RCI3, encodes a peroxidase that constitutes a component for stress tolerance. *Plant J.* **2002**, *32*, 13–24. [[CrossRef](#)]
55. Shen, Y.; Zhang, R.; Xu, L.; Wan, Q.; Zhu, J.; Gu, J.; Huang, Z.; Ma, W.; Shen, M.; Ding, F.; et al. Microarray Analysis of Gene Expression Provides New Insights Into Denervation-Induced Skeletal Muscle Atrophy. *Front. Physiol.* **2019**, *10*, 1298. [[CrossRef](#)]
56. Missirian, V.; Conklin, P.A.; Culligan, K.M.; Huefner, N.D.; Britt, A.B. High atomic weight, high-energy radiation (HZE) induces transcriptional responses shared with conventional stresses in addition to a core “DSB” response specific to clastogenic treatments. *Front. Plant Sci.* **2014**, *5*. [[CrossRef](#)]
57. Zhou, M.; Callahan, J.B.; Reyes, M.; Stasiak, M.; Riva, A.; Zupanska, A.K.; Dixon, M.A.; Paul, A.-L.; Ferl, R.J. Dissecting Low Atmospheric Pressure Stress: Transcriptome Responses to the Components of Hypobaric Stress in Arabidopsis. *Front. Plant Sci.* **2017**, *8*. [[CrossRef](#)] [[PubMed](#)]
58. Zhao, C.; Zhang, H.; Song, C.; Zhu, J.K.; Shabala, S. Mechanisms of plant responses and adaptation to soil salinity. *Innovation* **2020**, *1*, 100017. [[CrossRef](#)]
59. Zhu, L.; Zhang, Y.-H.; Su, F.; Chen, L.; Huang, T.; Cai, Y.-D. A Shortest-Path-Based Method for the Analysis and Prediction of Fruit-Related Genes in Arabidopsis thaliana. *PLoS ONE* **2016**, *11*, e0159519. [[CrossRef](#)] [[PubMed](#)]