

Article

Network Analysis of Gene Transcriptions of *Arabidopsis thaliana* in Spaceflight Microgravity

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

¹ Department of Electrical and Computer Engineering, University of Puerto Rico, Mayaguez, PR 00681-9000, USA; Jairo.orocho@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.agrinoni@upr.edu (C.A.)

* Correspondence: vidya.manian@upr.edu; Tel.: +78-7832-4040 ext. 3094

Abstract: The transcriptomic datasets of the plant model organism *Arabidopsis thaliana* grown in the International Space Station provided by GeneLab have been mined to isolate the impact of spaceflight microgravity on gene expressions related to root growth. A set of computational tools is used to identify the hub genes that respond differently in spaceflight with controlled lighting compared to on the ground. These computational tools based on graph-theoretic approaches are used to infer gene regulatory networks from the transcriptomic datasets. The three main algorithms used for network analyses are LASSO, Pearson correlation, and the HITS algorithm. Graph-based spectral analyses reveal distinct properties of the spaceflight microgravity networks for the WS, Col-0, and mutant *phyD* ecotypes. The set of hub genes that are significantly altered in spaceflight microgravity are mainly involved in cell wall synthesis, protein transport, response to auxin, stress responses, and catabolic processes. Network analysis highlights five important root growth-regulating hub genes that have the highest outdegree distribution in spaceflight microgravity networks. These concerned genes coding for proteins are identified from the Gene Regulatory Networks (GRNs) corresponding to spaceflight total light environment. Furthermore, network analysis uncovers genes that encode nucleotide-diphospho-sugar interconversion enzymes that have higher transcriptional regulation in spaceflight microgravity and are involved in cell wall biosynthesis.

Keywords: gene expression; graphs; network analysis; root growth; cell wall biosynthesis; spaceflight microgravity; *Arabidopsis thaliana*

Citation: Manian, V.; Orozco, J.; Gangapuram, H.; Janwa, H.; Agrinoni, C. Network Analysis of Gene Transcriptions of *Arabidopsis thaliana* in Spaceflight Microgravity. *Genes* **2021**, *12*, 337. <https://doi.org/10.3390/genes12030337>

Academic Editor: Bernd Weisshaar

Received: 26 December 2020

Accepted: 23 February 2021

Published: 25 February 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 (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Gravity plays a key role in plant growth [1]. *Arabidopsis thaliana* (*Arabidopsis*) is a flowering plant that has been chosen by biologists as a model plant organism to study the effects of gravity and other environmental stressors, since it has a short life cycle and because of the existence of a multitude of mutants and transgenic plants. The current visibility of *Arabidopsis* research reflects the growing realization among biologists that this simple angiosperm can serve as a convenient model not only for plant biology but also for addressing fundamental questions of biological structure and function common to all Eukaryotes [2]. To achieve the goal of providing long term life support for deep space missions a fundamental knowledge on plant growth in spaceflight has to be gained. The International Space Station (ISS) is currently equipped with all capabilities to collect data needed to address fundamental questions of plant physiology and development in spaceflight microgravity [3].

In this article, we are interested in identifying the genes that show transcriptional abundance under the influence of spaceflight microgravity, and the genes primarily influencing root growth and cell wall biosynthesis in spaceflight. The transcriptomic profile of *Arabidopsis* has been studied in simulated microgravity [4]. The effect of gravity and mechanical stimulation on the growth of the root apex of *Arabidopsis* has resulted in transient changes in gene expressions [5]. The effect of spaceflight microgravity on *Arabidopsis* root growth is presented in [6]. Cell wall modeling in the roots, shoots, and hypocotyls of plants is an important metabolic adaptation brought about by genes and proteins. Metabolic phenotyping of *Arabidopsis* has been carried out on the ground [7]. Metabolic fingerprinting is important for improving the quality of crops grown in spaceflight. Several of the cell wall proteome changes have been corroborated with transcriptomic changes in spaceflight [8]. As plants are important for the bioregenerative life support in spaceflight, the cell wall perturbations may affect the yield of crops grown in the altered spaceflight environment [9]. Of the 33 XTH *Arabidopsis* root growth genes, XTH17 to XTH20 are the main contributors [10]. Nucleotide sugar interconversions in plants are discussed in [11,12]. From terrestrial studies, nucleotide-diphospho-(NPD) sugar interconversion enzymes (NSE) of the small dehydrogenase/reductase superfamily are given in [13]. Plant Cell Wall (PCW) genes were identified using gene co-expression analysis in [14], and it was shown how the co-expression modules are transcriptionally co-regulated [15]. These sugars require subtle quantitative control during plant root growth and development. Hence, it is important to identify the genes that encode these enzymes involved in regulating growth of adventitious roots, and their waving and skewing in spaceflight microgravity.

Researchers have used various ecotypes and mutants of *Arabidopsis* to isolate genes with transcriptional variation in spaceflight environment compared to on the ground. Since we are not doing the experiment and performing the data collection, we rely on datasets available from GeneLab for plants grown in spaceflight microgravity. GeneLab provides transcriptomic datasets for various ecotypes of *Arabidopsis* from spaceflight. The *Arabidopsis* ecotypes Col-0, Ws-2, Ler-0, and Cvi-0 showed differential gene expression due to oxidative stress in spaceflight [16]. Differential gene expression analysis of *Arabidopsis* exposed over six days to blue light at different levels of gravity in comparison to unit gravity control has been presented in [17], and it was found that blue-light phototropism may be enough to reduce stress due to lower levels of gravity in spaceflight. It was found that plants in spaceflight microgravity grow slower than plants in unit gravity on the ground. Two ecotypes of *Arabidopsis* (WS and Col-0) were grown in spaceflight under directional light, and it was found that skewing and waving of roots were independent of gravity, the roots were negatively phototropic, and both ecotypes skewed differently in the presence of light, due to a lack of gravity [18]. Major organ-specific and cellular remodeling of *Arabidopsis* has been identified from its growth in ISS and physiological adaptation to spaceflight [3] [19]. Previous literature, cited above, characterized unique transcriptions in WS, Col-0, and PhyD mutant of *Arabidopsis* using heat maps and fold change analysis methods. For all the genome analysis done, no phenotypic measurements have been presented or discussed. However, the gradient-light experiments conducted in the ISS resulted in the plant root growth in a skewing pattern. These were dramatic observations of phenotype changes concerning root growth and cell wall response in microgravity. Hence, we have chosen the GLDS-7 [3] and GLDS-120 [20] datasets collected under different lighting settings for applying our network analysis methods.

The biochemical interactions within a cell of a biological system are formulated as gene regulatory networks for modeling biological processes. GRNs are used to represent regulatory and transcriptional functions that underlie cellular, molecular, and biological phenomena. GRNs have become an indispensable tool that facilitates interpretation of biological processes, molecular functions, protein interactions, genome to phenome analysis, cellular differentiation, identification of genes as biomarkers, and the response of an organism to external stimuli. A Bayesian network generation method for gene regulatory

network constructed for Arabidopsis has been presented [21]. Several machine learning algorithms have been implemented for the generation of these networks [22]. In [23], several statistical and machine learning methods have been reviewed for the prediction of GRNs from transcriptome datasets. In [24], many algorithms for GRN inferencing from single-cell transcriptomic data are benchmarked using simulated data. Several GRNs have been generated to study root development in Arabidopsis [25]. Vascular root development has been represented using GRNs in [26]. Furthermore, Arabidopsis secondary cell wall synthesis has been studied using a protein-DNA regulatory network and genes involved in secondary cell wall modeling, which form the cell shape and the bulk of the plant biomass [27]. Other GRNs for Arabidopsis growth and development are presented in [28]. Gene Ontology is the process of classifying an annotated gene to a particular function. Gene set enrichment analysis is used to determine changes of expression in groups of a priori defined gene sets. Additional web-tools and databases are available for Arabidopsis gene expression analysis [29]. These GRN studies do not identify hub genes and the target (authority) genes they regulate in spaceflight microgravity compared to on the ground for important biological processes that are affected in spaceflight, nor perform rigorous graph theoretic network analysis and derive biological interpretations.

We have used three basic graph-based network approaches in this paper (1) the Least Absolute Shrinkage and Selection Operator (LASSO), (2) Pearson correlation, and (3) Hypertext Induced Topic Search (HITS) methods. LASSO and Pearson correlation methods are applied to construct GRNs of transcriptional regulations in spaceflight microgravity as well as ground control. Section 2 presents the materials and methods used for GRN inferencing and analysis. In Section 3, we present the results of the GRN analysis of the significant hub genes (identified from Pearson and HITS method) with the highest transcriptional abundance in spaceflight microgravity. Results from network analysis of the Arabidopsis ecotypes and genotype are discussed in this section. The root growth and cell wall genes affected in spaceflight microgravity are isolated and their molecular functions are discussed. Section 4 presents the conclusions.

2. Materials and Methods

2.1. GeneLab Arabidopsis Datasets

We have mined GeneLab GLDS-7 and GLDS -120 datasets for identifying hub genes that are affected by spaceflight. The GLDS-7 dataset contains microarray gene expression data for the Arabidopsis WS and Col-0 ecotypes for different tissues (roots, leaves, and hypocotyls) for the plants grown in ground control and spaceflight in white light. GLDS-120 contains the transcriptomics data of the Arabidopsis Wassilewskija (WS) and Columbia (Col-0) ecotypes, and the gene *Phytochrome D* (*phyD*) mutant of ecotype Col-0. The different stressors imposed on the plants are 4–6 $\mu\text{moles m}^{-2} \text{s}^{-1}$ of total light and dark in spaceflight and on the ground. The design of the experiment was to investigate the effect of light on Arabidopsis root growth when the gravity stimuli were nullified. The *phyD* mutant was used, since it is a known light-related mutant in Col-0 background. RNAseq is used for transcription profiling. Gene expression data are available for pairs of combinations of genotypes and ecotypes.

2.2. Graph-Based GRN Inferencing

We define a graph-based representation for the gene expressions. Formally, a graph is a pair of sets $G = (V, E)$ where V is the vertices (molecules, genes, proteins, nodes, points) and E is an ordered pair of vertices (u, v) that are called the edges (arcs), respectively [30]. One important advantage of the representation of networks by graphs is that they have an intuitive graphical representation. In a directed graph $G = (V, E, o, t)$, $e = (u, v)$ are the edges, the origin of e is denoted by $o(e)$, and the terminal v is denoted by $t(v)$. A graph is weighted if each edge e is assigned a positive real-valued weight $w(e)$. Principal Component Analysis (PCA) method is applied to de-correlate the gene expression values. PCA is

done to reduce the dimensions to three main components corresponding to the largest eigenvalues of the covariance matrix computed from the graph representation of the gene expression data. PCA projects the data on to an orthogonal space, such that the variance of the projected data is maximized. This space is lower in dimension with minimum projection cost. PCA is applied to the gene expression datasets before performing both LASSO regression and Pearson correlation GRN construction.

2.2.1. LASSO Regression

Lasso is a linear regression algorithm that shrinks the data towards a central point. Lasso regression mostly leads to sparse models. Lasso uses the l_1 regularization, which means it adds a penalty that is equal to the sum of the absolute value of the coefficients. Lasso regression is computationally feasible and can fit any kind of statistical data [31]. l_1 regularization is applied to eliminate the relationship between the genes with the least importance. The Lasso regression is constructed according to the description in [31]. We have used the SILGGM package in R for the Lasso Regression [32]. In this algorithm, every gene present in the data is considered as a transcription factor, as well as a target gene, and the correlation between each gene with the other is calculated using the scaled Lasso algorithm. The results from the regression are used to construct GRNs in Cytoscape for the sub-modules of the genes of Arabidopsis spaceflight and ground datasets. The method has been applied to the GeneLab GLDS-120 dataset.

2.2.2. Pearson Correlation

Pearson's correlation coefficient between two genes (variables, nodes) is defined as the covariance of the two variables divided by the product of their standard deviations:

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} \quad (1)$$

Pearson's correlation coefficient is a test that measures the statistical relationship between two random variables. If the association between the elements is not linear, then the coefficient is not specifically identified. In this study, we use the Pearson correlation to construct the co-expressed gene network to identify genes with correlated expression patterns. This method is based on an adjusted threshold p -value to accommodate multiple hypotheses; the correlation between expression patterns is calculated for every pair of genes. The Pearson correlation coefficient is used as the measure of correlation (between -1 to 1) and gene pairs are considered to be co-expressed if their p -value is less than or equal to the corrected threshold p -value. The t -test is used to establish if the correlation coefficient is significantly different from zero, and hence, that there is evidence of an association between the two variables. The genes are treated as nodes and the relation between them is an edge. Therefore, the nodes with the highest degree are related to a higher number of target or authority genes. Those genes that have the highest degree or intra-module (within graph) connectivity are treated as hub genes to construct the gene regulatory networks. Figure 1 shows the steps involved in generating the GRN from the microarray gene expression dataset using this method.

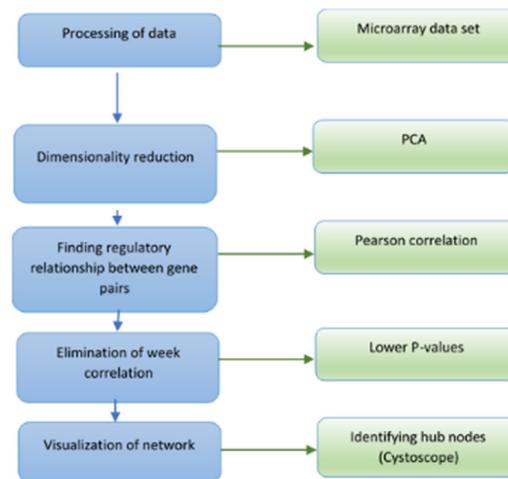


Figure 1. Flow chart of the procedure for gene regulatory network inferencing using the graph-based Pearson correlation method.

2.2.3. Sparse Networks and the Weighted HITS Algorithm

We show that the GRN generated for the GeneLab Arabidopsis datasets are sparse networks or graphs. We define sparsity precisely so that we utilize quantitatively time and speedups in HITS and other algorithms. Briefly, we recall the following definitions obtained from [30] and [33].

Directed graph: The graph (V, E, o, t) is said to be directed if directed edges are allowed, i.e., not all edges have reverse edges as members of E . In a directed graph, $G=(V,E,o,t)$, the edges $e=(u,v)$, the origin of e is denoted by o , and the terminal v is denoted by $t(v)$.

Bipartite graph: The graph is called *bipartite* if $V = V_1 \cup V_2$ such that edges connect node V_1 to V_2 . The expanding constant or isoperimetric constant (or Cheeger constant) of X is defined as:

$$h(X) = \min_{\emptyset \neq F \subset V} \frac{|\partial F|}{\min\{|F|, |V \setminus F|\}} \quad (2)$$

where $F \subset V$ and the boundary ∂F is the set of edges connecting F to $V \setminus F$. The isoperimetric constant is difficult to compute, but we can bound its value using the spectral gap. The Spectral Gap is defined as the difference between the two highest eigenvalues of the adjacency matrix of the graph.

Sparse graphs: We say that a graph (or the corresponding network) is sparse if $h(X)$ is large. One of the major results of Cheeger for Riemannian manifolds and its equivalent for graphs is that a large spectral gap implies large $h(X)$. Computing $h(X)$ is an NP-Complete problem, while the computing spectral gap is at most cubic and many times linear; we work with the spectral gap here.

Spectral gap: For a graph X , the Laplacian eigenvalues can be ordered as $|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_n|$ (X may be directed or undirected, weighted or unweighted, simple or not). The spectral gap is defined as: $\delta_\lambda = |\lambda_1| - |\lambda_2|$. By normalizing the Laplacian matrix of X , the eigenvalues are $1 = \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n > 0$, and the Laplacian spectral gap will be: $\delta_\lambda = 1 - |\lambda_2|$. The spectral gap is also known as a random walk, in terms of this concept λ_2 is the most important eigenvalue. Note that if the spectral gap is 0, $\lambda_2 = 1$ (Γ is not (strongly) connected or Γ is bipartite), and a typical random walk will not converge to a unique distribution or dominant eigenvector. As long as the spectral gap is greater than 0, which means $|\lambda_2| < 1$, then the random walk converges to a unique dominant eigenvector, and

the spectral gap measures the rate of convergence; the larger the spectral gap (the smaller $|\lambda_2|$), the better the network flow (large $h(X)$, diffusion, mixing, random walk, expansion, sparsity, and other highly desirable properties of the network X).

Girth: The girth of X is the smallest positive integer r such that $\text{Trace}(A^r) > 0$. Let $d = d(X)$ be the smallest integer (if it exists), so that for every pair of vertices (u, v) , there is a walk of length at most d from u to v . Then, $d(X)$ is called the diameter or maximum eccentricity of the graph X .

Clustering coefficient: The clustering coefficient models the degree of clustering of a subset of nodes. A node is selected, and we see how connected the node is with other nodes that are also connected to it. The clustering coefficient is defined by:

$$C_i = \frac{L_i}{K_i(K_i - 1)/2} \quad (3)$$

where L_i is the number of connections between K_i , which are the immediate neighbors of i . The clustering coefficient is used to characterize network modularity, which is a strength of measure of a network division into modules or groups.

Degree distribution: The degree distribution is the number of neighbors connected to a node; in other words, it is the number of edges incident on a node. The degree distribution can give information about the structure of a network. The networks can be directed or undirected. In the undirected case, the degree of a node i is the number of connections it has, and it can be represented as an adjacency matrix:

$$K_i = \sum_j a_{ij} \quad (4)$$

where the sum is over all the nodes. For directed graphs, there are two types of degree distributions: in-degree, which is the number of connections entering the node, and out-degree, which is the number of outgoing connections. For this case, the sum of the i th rows and i th columns of the adjacency matrix are:

$$K_i^{total} = \sum_j a_{ij} + \sum_j a_{ji} \quad (5)$$

The degree distribution for a directed graph is the fraction of vertices of degree k *in* and k *out*.

Biomolecular networks such as GRNs 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 [34], called the HITS algorithm. Originally, it was meant for networks such as the Internet; however, now, it is used to study a variety of biomolecular networks, such as GRNs and Protein–Protein Interaction (PPI) networks [35,36]. 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 for 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. The weighted HITS algorithm computes the hub \vec{h}_k (resp. authority \vec{a}_k) nodes iteratively as: $\vec{h}_k = \psi_k \phi_{k-1} A A^T \vec{h}_{k-2}$ and $\vec{a}_k = \phi_k \psi_{k-1} A^T A \vec{a}_{k-2}$. These iterations converge to the dominant eigenvector of the real symmetric matrices $A A^T$ ($A^T A$). These give us asymptotically hub and authority weights. A version of this algorithm is implemented and applied to the networks in the software SAGE. 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 spectral gap analysis shows that the graphs are sparse but highly connected, helping us carry out HITS with a high degree of iterations. The weighted HITS

algorithm has yielded important information about biomolecular networks [37–45]. For network topology, we refer to [46], and for the latest on the origin of biomolecular networks, topological, combinatorial, and spectral methods, we refer to [47].

3. Results and Discussion

The Pearson correlation method is used to generate weighted directed graphs. Two columns of 20,000 samples of gene expression values of *Arabidopsis* ecotypes WS and Col-0 root growth in spaceflight from GLDS-7 and GLDS-120 datasets are plotted in Figure 2 (A) and (B), respectively. X-axis is the first column of the gene expression values and y-axis is the second column of the gene expression values. The plot shows that the values are highly correlated. Hence, PCA is first applied to reduce the dimensionality of the gene expression data to three dimensions. The three PCA components are then used for computing Pearson correlation and generating the network, as explained in Section 2.2.2. All the gene expression values from GLDS-7 and GLDS-120 spaceflight and ground data for *Arabidopsis* WS ecotype root growth are used for generating the networks. In Table 1, sample correlation values between 13 genes, their t-statistics, and their *p*-values are given. Correlation values range from -1 to 1 . *P*-values can also be used to construct GRNs by thresholding them between, for example, 0.01 , and 0.0001 , which will give the same network as a correlation threshold of 0.05 . The threshold can be adjusted depending on the required level of complexity of the inferred GRN. Figures 3 and 4 show sub-graphs extracted from GLDS-7 root growth spaceflight and ground gene expression data, respectively, with a correlation threshold of 0.05 . This threshold selects 11.3% of genes from the GLDS-7 dataset and 28.7% genes from the GLDS-120 dataset. These are weighted and directed sub-graphs extracted from the GLDS-7 spaceflight and ground GRNs, respectively. The weights are the correlation values (positive and negative) between the hub gene and the target gene, shown in Figures 3 and 4. Negative interactions are important because they show an activation–inhibition two-way expression pattern. While positive correlations indicate shifting, scaling, or geometric pattern in gene expression data, negative correlations indicate opposite and complementary expression patterns, and self-negative regulation of transcription factors (TFs). Hence, we have extracted both maximal positive and negative correlations for constructing the GRNs. These gene interactions have not been validated by follow up experiments; however, they have been constructed from gene expression data obtained originally from experiments conducted in the ISS.

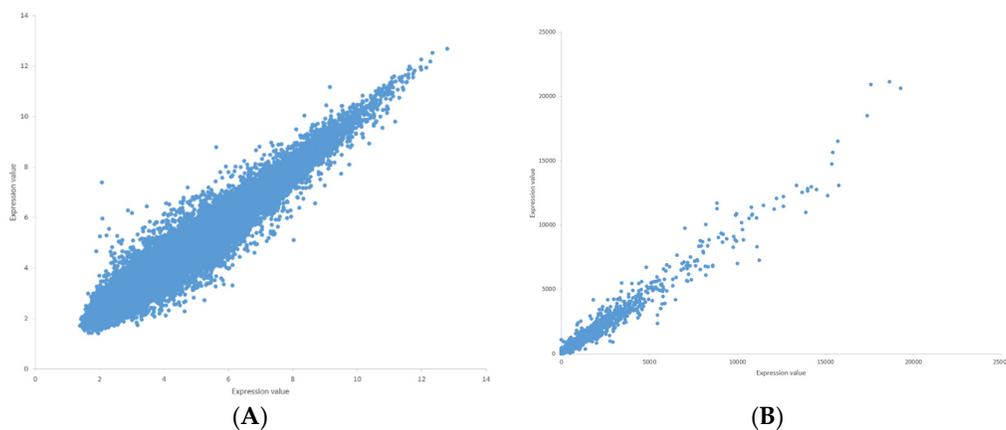


Figure 2. (A) Plot of gene expression values for *Arabidopsis thaliana* WS root growth in spaceflight microgravity from GLDS-7 dataset. The plot shows high correlation between the first column (*x*-axis) and second column (*y*-axis) of gene expression values for 20,000 genes. (B) Plot of gene expression values for *Arabidopsis thaliana* Col-0 root growth in spaceflight microgravity from GLDS-120 dataset. The plot shows high correlation between the first column (*x*-axis) and second column (*y*-axis) of gene expression values for 20,000 genes.

Table 1. Hub genes and target authority genes for 13 hub genes with their correlation value, t-statistics, and *p*-value.

Interactor Gene A	Interactor Gene B	Correlation	t-Statistics	p-Value
AT1G03430	AT1G11185	-0.95787	-4.716511	0.04213
AT1G03430	AT1G12390	0.984801	8.018483	0.0152
AT1G03430	AT1G17170	-0.98068	-7.090435	0.01932
AT1G03430	AT1G17180	0.937057	3.795227	0.06294
AT1G03430	AT1G21140	0.454929	0.722456	0.54507
AT1G03430	AT1G24240	0.993113	11.98728	0.00689
AT1G03430	AT1G26470	0.98765	8.914699	0.01235
AT1G03430	AT1G27130 AT1G27140	0.998482	25.63509	0.00152
AT1G03430	AT1G27570	-0.95498	-4.552277	0.04502
AT1G03430	AT1G30360	0.9732	5.985038	0.0268
AT1G03430	AT1G32460	0.4588	0.730233	0.5412
AT1G03430	AT1G34180	-0.99271	-11.64964	0.00729
AT1G03430	AT1G34844	0.947816	4.204318	0.05218

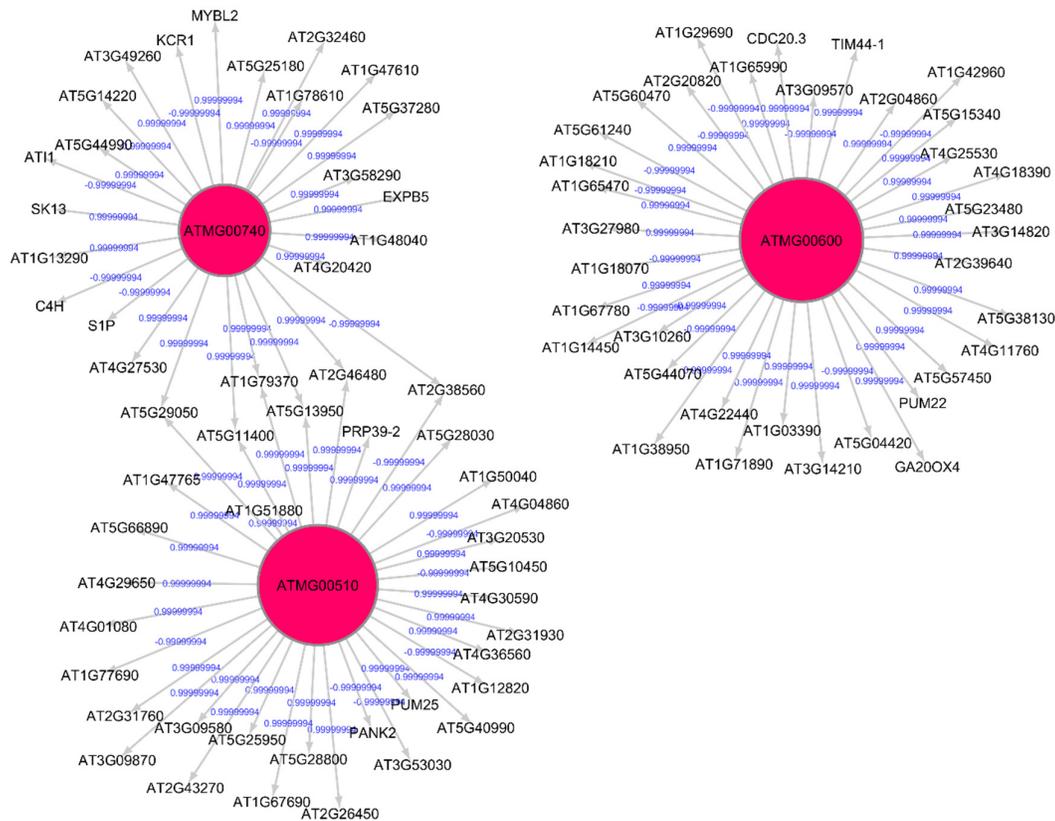


Figure 3. Gene regulatory sub-graph for *Arabidopsis thaliana* root growth in spaceflight microgravity. The hub genes are the red circles representing the strong interactions with differentially expressed genes. The target (authority) genes are shown without circles. The weights on the arrows are the correlation values between the hub gene and the target gene.

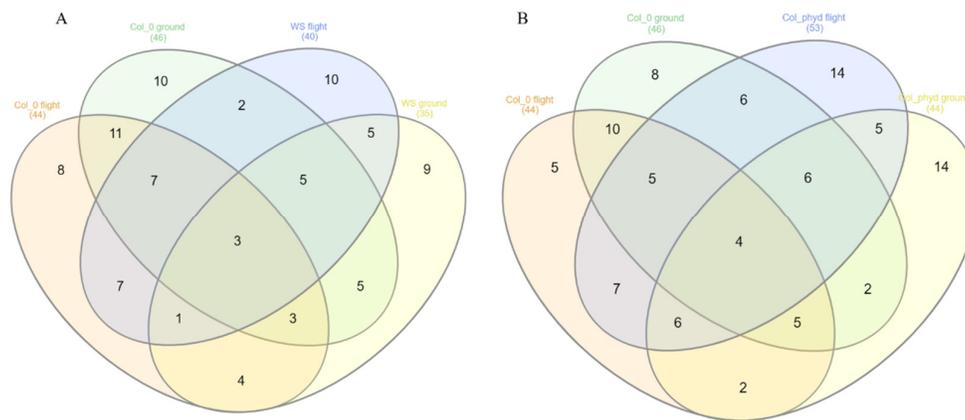


Figure 5. Venn diagrams showing the number of shared hub genes from the gene regulatory networks for *Arabidopsis thaliana*. (A) Between the WS and Col-0 ecotypes in spaceflight and ground control for both total light and dark environments. (B) Between Col-0 ecotype and Col-0 mutant *phyD* in spaceflight and ground control for both total light and dark environments.

Figure 6 shows the Pearson correlation gene regulatory sub-graphs extracted from GRNs of the GLDS-120 dataset for the catabolic process for the *Arabidopsis* WS ecotype root growth in spaceflight light environment. The number of hub genes in the GRN representation can be controlled by suitably adjusting the threshold value. The lower the threshold, the higher the number of hub genes can be detected and vice versa. In all our networks, we have used a nominal correlation threshold value of 0.05. The original networks have more than 600 genes involved in regulatory pathways. The GSEA of these hub genes present in spaceflight microgravity and ground control GRNs for WS, Col-0, and Col mutant *phyD* ecotypes grown in different light environments are presented below.

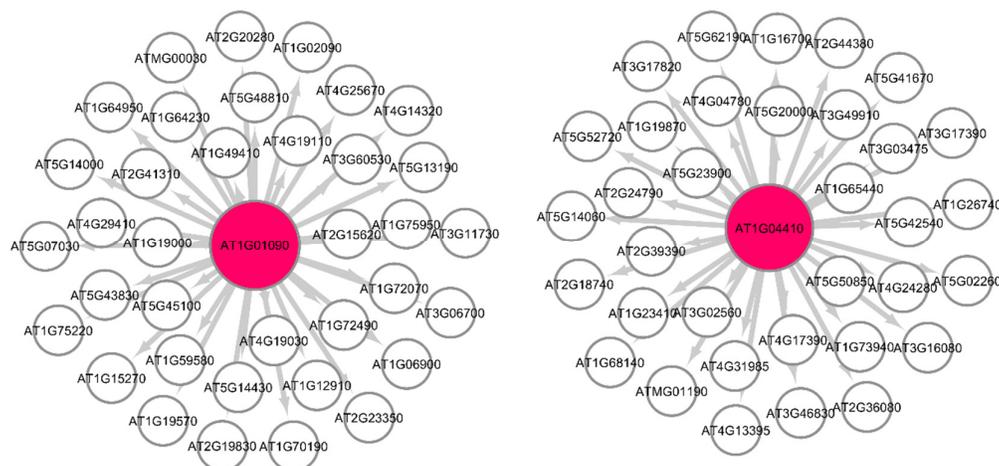


Figure 6. Gene regulatory sub-graph for the catabolic process in *Arabidopsis thaliana* Col-0 mutant *phyD* in spaceflight total light environment. The hub genes are the red circles representing the strong interactions with differentially expressed genes. The target (authority) genes are white circles.

3.1. Gene Set Enrichment Analysis (GSEA)

The hub genes, shared genes for spaceflight and ground environments, and their gene ontology from the GeneLab GLDS-7 root growth datasets are given in Table 2. The ShinyGO tool (<http://bioinformatics.sdstate.edu/go/>) is used to perform GSEA of the hub

genes for the GeneLab GLDS-7 root growth GRN. The hub genes for spaceflight microgravity and ground overlap considerably and are involved in molecular functions, biological processes, and oxidation-reduction process. They enable NADH dehydrogenase activity, and are involved in NAD and quinone binding. Contrary to GLDS-7 GRN functions, the GRNs constructed for the Arabidopsis WS, Col-0, and Col mutant *phyD* show significantly different regulations in spaceflight in total light. These GRNs do share genes with ground control GRNs, but have other genes that show significant transcriptions in spaceflight microgravity, and these genes are not shared by the ground GRNs. Hence, these genes can be considered to be affected by spaceflight microgravity alone. These genes are related to biological processes such as DNA integration, the RNA-dependent DNA biosynthetic process, and the DNA metabolic process. We performed a GSEA analysis comparing hub genes for the WS wild ecotype and Col-0 ecotype. The hub genes for the WS ecotype are associated with biological processes such as the catabolic process, response to abiotic stimulus, localization, multicellular organismal process, developmental process, and anatomical structure development. The hub genes for the Col-0 ecotype perform similar functions to WS ecotype such as the catabolic process, regulation of metabolic process, positive regulation of biological process, cellular response to stimulus, signaling, multicellular organismal process, developmental process, and anatomical structure development. The shared hub genes between the Col-0 ecotype and Col-0 mutant *phyD* perform functions such as regulation of metabolic process, multicellular organismal process, developmental process, positive regulation of biological process, anatomical structure development, and catabolic process. The non-shared hub genes of *phyD* in spaceflight microgravity perform significant functions such as the cellular component organization of biogenesis, response to abiotic stress, and response to external stimulus.

Table 2. Hub genes from the Pearson Correlation GRN and HITS method and shared hub genes for GLDS 7 for *Arabidopsis thaliana* root growth in spaceflight and ground control with their gene ontology.

Arabidopsis Root	ATG Number	Enriched Gene Set (Biological Process)
Root growth spaceflight hub genes	ATMG00880 ATMG00840 ATMG00660	Part of mitochondrion; Enables molecular function; Involved in biological process
Root growth ground control hub genes	ATMG00720 ATMG01040 ATMG01020 ATMG00580	Part of mitochondrion Enables molecular function; Involved in biological process; ATP synthesis coupled electron transport; Involved in oxidation-reduction process
Shared hub genes for root growth in spaceflight and ground control	ATMG00860 ATMG00510 ATMG00750 ATMG00740 ATMG00720 ATMG00600 ATMG00680 ATMG00820	Part of mitochondrion Enables molecular function; Involved in biological process; Enables NADH dehydrogenase activity; Part of mitochondrial respiratory chain complex 1; Involved in oxidation-reduction process; Enables NAD binding; quinone binding

3.1.1. GSEA for Common Hub Genes

Eight common hub genes between the spaceflight and ground environments have been identified, as shown in Table 2. Using g:Profiler, we found that these genes, including the gene ATMG00600 (ORF 106C), are uncharacterized mitochondrial protein and their

function is DNA integration. AT1G04080 is a shared gene between the Col-0 and WS ecotypes, which encodes a U1 small nuclear ribonucleoprotein (snRNP) factor involved in alternative splicing. AT1G02930 and AT1G04980 are the shared genes between the Col-0 and Col-0 mutant *phyD*, which are involved in the defense response to bacterium, glutathione metabolic process, response to cadmium ion, response to oxidative stress, water deprivation, heat, response to unfolded protein, and ubiquitin pathway. The Pearson and LASSO methods have led to the identification of signature sub-networks that show the interaction across several important metabolic and signaling processes, and Transcription Factors (TFs) for Arabidopsis growth in spaceflight microgravity. Both methods were able to identify the transcription factors AT1G06040, AT1G01720, and AT1G04850. AT1G06040 encodes salt tolerance protein and AT1G01720 is involved in the cellular response to hypoxia and has a DNA binding TF activity. The TF AT1G04850 is a ubiquitin-associated (UBA)/TS-N domain-containing protein, and its function is zinc ion binding. From the GRNs, we can identify genes with the maximum degree (hub genes) that regulate a larger number of authority (target) genes.

3.1.2. Gene Ontology of Transcription Factors (Genes) in Spaceflight

Thirty-eight hub genes with transcriptional abundance in spaceflight have been identified from the total light environment gene expression datasets of GLDS-120. Their ATG number, gene name, and gene ontology are presented in Supplementary Table S1. These are hub genes, as their transcriptional responses are abundant only in spaceflight microgravity network and are not found to be hub genes in the corresponding terrestrial GRNs. Hence, these genes are functioning only in spaceflight microgravity conditions. From their ontology, it is evident that they are responsible for important functions necessary for the plant's growth in an altered environmental setup compared to the terrestrial environment. The plant can sense the changes in abiotic stressors such as gravity, oxygen, and radiation and can step up transcriptional regulation to adapt to the new environment successfully. The transcription functions of these genes are chiefly cell wall biosynthesis, protein transport, stress responses, catabolic processes, auxin, and protein regulation. It is also significant that these genes were identifiable only from the GLDS-120 dataset, which had a different lighting environment in spaceflight. Through novel graph theoretic methods applied to GRNs of Arabidopsis, we can explain phenomenological traits that have been observed in spaceflight vis-à-vis on the ground.

3.2. Topological and Spectral Analyses of Gene Regulatory Networks

Topological and algebraic analyses of the GRNs are given in Table 3. The WS ecotype in total light spaceflight environment has the largest spectral gap, which indicates that this network has the highest performance with stable synchronization, rapid convergence, mixing, expansion, and diffusion [47]. Random walks propagate more easily in networks with a large spectral gap. The large spectral gap has also made several algorithms we use tractable for such massively large GRN and indicates that in technical terms, these GRNs are highly sparse. This corroborates with results from other studies that have shown that the WS genotype increases the transcriptional cost of spaceflight adaptation [20]. The diameter of this network is 6, indicating that it has longer gene signaling pathways, and the hub genes of this network regulate about 1000 target genes. The WS and Col-0 ecotypes in the ground light environment show topological similarities with the same shortest path length, and the two networks show regulation of about 45 target genes.

Table 3. Topological and spectral analysis for Arabidopsis (AT) spaceflight and ground gene regulatory networks.

Dataset	Spectral Gap	Average Clustering Coefficient	Average Shortest Path Length (Distance)	Diameter (Maximum Eccentricity)	Authority Genes (Outdegree)
AT root spaceflight	39.228	0.00897	1.452	3	95
AT root ground	59.96	0.00315	1.2017	3	166
AT WS spaceflight light	72.9	0.01272	1.89	6	1237.04
AT WS ground light	0.45	0.01788	1.11	2	45.63
AT WS spaceflight dark	3.99	0.01996	1.069	2	50.41
AT WS ground dark	2.32	0.02011	1.069	3	43.36
AT <i>phyD</i> spaceflight light	0.79	0.01000	1.036	2	35
AT <i>phyD</i> ground light	0.22	0.01996	1.069	2	50.41
AT <i>phyD</i> spaceflight dark	0.001	0.01219	1.146	2	38.24
AT <i>phyD</i> ground dark	3.15	0.0219	1.125	2	53.29
AT Col-0 spaceflight light	0.13	0.0212	1.107	2	40.26
AT Col-0 ground light	4.00	0.0146	1.113	2	44.28
AT Col-0 spaceflight dark	0.20	0.0224	1.066	1	29.09
AT Col-0 ground dark	2.67	0.0187	1.076	1	33.42

We show, with our spectral analysis methods, through the spectral estimates on the Cheeger constant that our networks are sparse, that the HITS algorithm works much faster, and we have been able to identify hubs and authority nodes (genes) in the networks. Rather than just the degree distribution, hubs and authority nodes are computed iteratively with newer algorithms considering weights using its implementation in the software SAGEMATH. Improved algorithms exist, e.g. [35], and some of Kleinberg's works.

3.3. Arabidopsis Root Growth and Cell Wall Biosynthesis in Spaceflight Microgravity

Out of the 33 root-growth related genes of Arabidopsis, we have identified five significant hub genes listed in Table 4 to have higher regulations in spaceflight. Figure 7 shows the subgraph that is extracted from the original Pearson correlation GRN by thresholding the genes with higher correlation values and involved in the cell wall biosynthesis metabolic process that relates the co-expression of these genes in spaceflight microgravity for the Arabidopsis Col mutant *phyD* ecotype. These five hub genes have higher regulations only in the spaceflight subgraph for the cell wall response process and not in the ground subgraph. These genes are over-represented in the spaceflight responsive transcriptome because they regulate plant growth in spaceflight microgravity and enable adaptation of the plant to the altered environment. Similarly, comparing the GRNs generated by the LASSO method for Arabidopsis WS ecotype in spaceflight and ground environments, the hub genes that are involved only in spaceflight for cell wall biosynthesis have been identified and listed in Supplementary Table S2. Figure 8 shows the subgraph for the cell wall biosynthesis process for Arabidopsis WS ecotype in spaceflight microgravity, which is also extracted similarly by thresholding the correlation values. These genes encode NSEs and have higher transcriptional regulation in spaceflight, as they were not found to be hubs in the GRNs for the terrestrial datasets. These enzymes catalyze a series of biochemical reactions and are involved in building cell wall polysaccharides. The origin and divergence of these enzymes is not known, and it is interesting to note that they are significantly transcribed to be indicative of higher sugar diversity in spaceflight microgravity. The higher regulations of these genes may be the reason for the phenotypic skewing or waving effect found in the roots, as well as other parts of the plant in spaceflight. The differential transcriptions of NSEs indicate that they regulate developmental, metabolic, and stress-related responses in the spaceflight microgravity environment.

Table 4. Root growth hub genes with transcriptional abundance in spaceflight microgravity and their most prominent gene ontology.

ATG Number	Hub Gene Name/Description	Enriched Gene Set (Biological Process)
AT2G36870	XTH32	Involved in cellular glucan metabolic process
AT2G18800	XTH21	Involved in cell wall organization
AT3G44990	XTH31	Involved in glucanase activity
AT3G23730	XTH16	Involved in the metabolic process
AT2G14620	XTH10	Involved in cellular glucan metabolic process

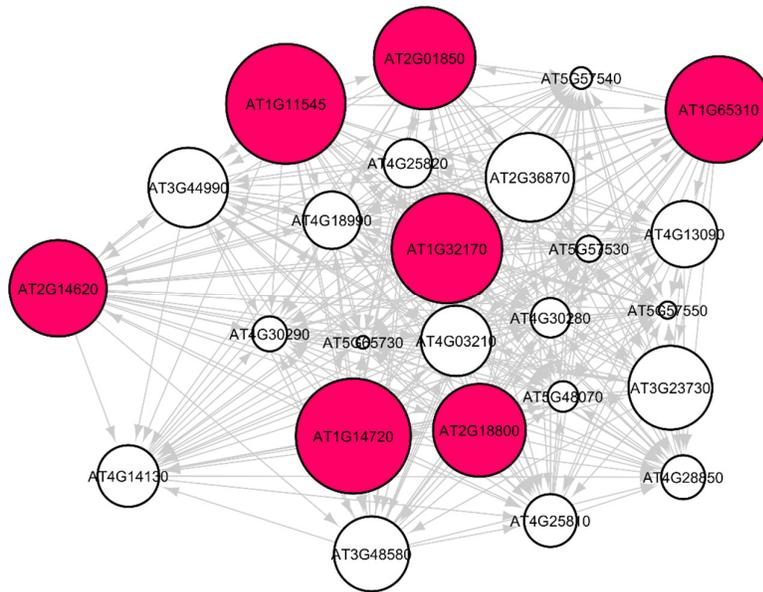


Figure 7. Gene regulatory sub-graph involving root growth genes in *Arabidopsis thaliana* Col-0 mutant *phyD* in spaceflight total light environment. The hub genes are the red circles representing the strong interactions with differentially expressed genes. The target (authority) genes are white circles. The hub genes that have a higher degree distribution are larger in size.

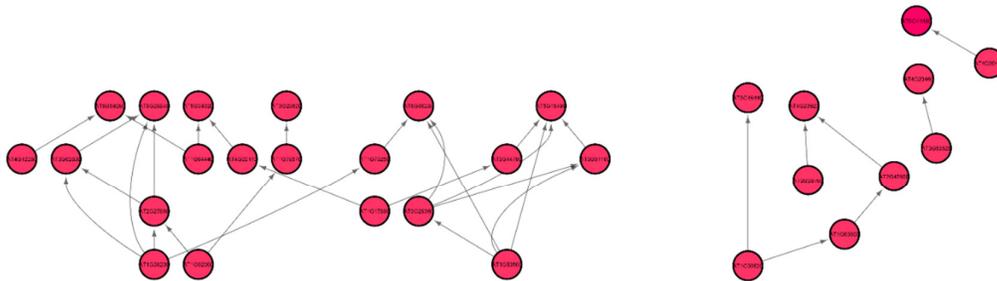


Figure 8. Gene regulatory sub-graph for the process of cell wall biosynthesis in *Arabidopsis thaliana* WS ecotype in spaceflight total light environment.

4. Conclusions

Graph-based network analysis methods and results have been presented for transcriptomic regulatory GeneLab *Arabidopsis* datasets. Previous studies were at the transcriptomic level. In this article, we have built gene regulatory networks for different ecotypes (via Pearson Correlation and LASSO methods), carried out both topological and algebraic network studies (the novel HITS algorithm for weighted and directed graphs to identify hubs and authority nodes), computed the spectral gap for the whole GRN, leading us to potential conclusions about phenomena such as diffusion and expansion of the GRN, and studied global and local topological parameters, such as clustering, average path length, eccentricity, and degree distribution. Therefore, our research is more comple-

mentary to prior studies and is more rigorous. Our novel graph-theoretic approach isolated hub genes regulated significantly different in spaceflight microgravity from terrestrial gene expressions. The GRN for the WS ecotype in spaceflight has the highest spectral gap, indicating that it has the highest transcriptional cost in spaceflight. GRN analysis showed that there are 38 genes with higher transcriptions in spaceflight microgravity and are involved in vital molecular and biological functions for the survival and growth of the plant in spaceflight.

While previous research has identified altered spaceflight transcriptions for cell wall remodeling, we have identified hub genes with higher transcriptional activity and that are involved in the cell wall response process only in spaceflight GRN; these hub genes are essentially metabolic phenotypes. The network analysis has detected xyloglucan endotrans glycosylase/hydrolases (XTHs) that are cell wall modifying enzymes mediating vital elongation growth of Arabidopsis in spaceflight. The XTHs act on the cell wall, making it more extensible, resulting in cellular expansion and the skewing phenomenon of the roots in spaceflight microgravity.

Through advanced network analysis, we can shorten the gap between genotype to phenotype analysis by identifying primary metabolites or metabolic phenotypes involved in the growth of plants in spaceflight for bioregenerative life support. In the absence of mass spectroscopy, hyperspectral imaging methods in the ISS, and lack of plant physiological measurements, genome to phenome network analysis offers an alternative for identifying primary metabolic phenotypes. The spaceflight response of these metabolites can be further correlated with phenotypic data extracted from images and metabolomics datasets acquired from experiments conducted in the ISS or from similar experiments on the ground. Phenotypic data will enable novel trait correlations employing genome to phenome network analysis.

Supplementary Materials: The following are available online at www.mdpi.com/2073-4425/12/3/337/s1, The genelab datasets are available at <http://www.genelab.nasa.gov/>. Supplementary Tables S1 and S2 are included in the Supplementary Materials.

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

Funding: This research was funded by NASA, grant number 80NSSC19M0167.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

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

Acknowledgments: The material contained in this document is based upon work supported by the National Aeronautics and Space Administration (NASA) grant 80NSSC19M0167. 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. Hoson, T. Plant growth and morphogenesis under different gravity conditions: Relevance to plant life in space. *Life* **2014**, *4*, 205–216, doi:10.3390/life4020205.
2. Koornneef, M.; Meinke, D. The development of Arabidopsis as a model plant. *Plant J.* **2010**, *61*, 909–921, doi:10.1111/j.1365-313X.2009.04086.x.
3. Paul, A.L.; Zupanska, A.K.; Schultz, E.R.; Ferl, R.J. Organ-specific remodeling of the Arabidopsis transcriptome in response to spaceflight. *BMC Plant Biol.* **2013**, *13*, 112, doi:10.1186/1471-2229-13-112.

4. Kamal, K.Y.; van Loon, J.J.W.A.; Medina, F.J.; Herranz, R. Differential transcriptional profile through cell cycle progression in Arabidopsis cultures under simulated microgravity. *Genomics* **2019**, *111*, 1956–1965, doi:10.1016/j.ygeno.2019.01.007.
5. Kimbrough, J.M. The fast and transient transcriptional network of gravity and mechanical stimulation in the Arabidopsis root Apex. *Plant Physiol.* **2004**, *136*, 1–16, doi:10.1104/pp.104.044594.
6. Ludvine, S.; Balzergue, S.; Carnero-Diaz, E.M.-M.M. *Action of Microgravity on Root Development*; GeneLab: 2012; Version 1.
7. Meyer, R.C.; Steinfath, M.; Lisec, J.; Becher, M.; Witucka-Wall, H.; Torjek, O.; Fiehn, O.; Eckardt, A.; Willmitzer, L.; Selbig, J.; et al. The metabolic signature related to high plant growth rate in Arabidopsis thaliana. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 4759–4764, doi:10.1073/pnas.0609709104.
8. Ferl, R.J.; Koh, J.; Denison, F.; Paul, A.L. Spaceflight induces specific alterations in the proteomes of Arabidopsis. *Astrobiology* **2015**, *15*, 32–56, doi:10.1089/ast.2014.1210.
9. Johnson, C.M.; Subramanian, A.; Pattathil, S.; Correll, M.J.; Kiss, J.Z. Comparative transcriptomics indicate changes in cell wall organization and stress response in seedlings during spaceflight. *Am. J. Bot.* **2017**, *104*, 1219–1231, doi:10.3732/ajb.1700079.
10. Somssich, M.; Khan, G.A.; Staffan, S.P. Cell wall heterogeneity in root development of Arabidopsis. *Front. Plant Sci.* **2016**, *7*, 1–11, doi:10.3389/fpls.2016.01242.
11. Yin, Y.; Huang, J.; Gu, X.; Bar-Peled, M.; Xu, Y. Evolution of plant nucleotide-sugar interconversion enzymes. *PLoS ONE* **2011**, *6*, e27995, doi:10.1371/journal.pone.0027995.
12. Reiter, W.; Vanzin, G.F. Molecular genetics of nucleotide sugar interconversion pathways in plants. *Plant Mol. Biol.* **2001**, *47*, 95–113.
13. Seifert, G.J. Nucleotide sugar interconversions and cell wall biosynthesis: How to bring the inside to the outside. *Curr. Opin. Plant Biol.* **2004**, *7*, 277–284, doi:10.1016/j.pbi.2004.03.004.
14. Cai, B.; Li, C.H.; Huang, J. Systematic identification of cell-wall related genes in populus based on analysis of functional modules in co-expression network. *PLoS ONE* **2014**, *9*, e95176, doi:10.1371/journal.pone.0095176.
15. Wang, S.; Yin, Y.; Ma, Q.; Tang, X.; Hao, D.; Xu, Y. Genome-scale identification of cell-wall related genes in Arabidopsis based on co-expression network analysis. *BMC Plant Biol.* **2012**, *12*, 1, doi:10.1186/1471-2229-12-138.
16. 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, doi:10.1002/ajb2.1223.
17. Herranz, R.; Vanderbrink, J.P.; Villacampa, A.; Manzano, A.; Poehلمان, W.L.; Feltus, F.A.; Kiss, J.Z.; Medina, F.J. RNAseq analysis of the response of Arabidopsis thaliana to fractional gravity under blue-light stimulation during spaceflight. *Front. Plant Sci.* **2019**, *10*, 1–11, doi:10.3389/fpls.2019.01529.
18. Paul, A.L.; Zupanska, A.K.; Ostrow, D.T.; Zhang, Y.; Sun, Y.; Li, J.L.; Shanker, S.; Farmerie, W.G.; Amalfitano, C.E.; Ferl, R.J. Spaceflight transcriptomes: Unique responses to a novel environment. *Astrobiology* **2012**, *12*, 40–56, doi:10.1089/ast.2011.0696.
19. Paul, A.L.; Zhou, M.; Callahan, J.; Reyes, M.; Stasiak, M.; Riva, A.; Zupanska, A.K.; Dixon, M.; Ferl, R. Patterns of Arabidopsis gene expression in the face of hypobaric stress. *AoB Plants* **2017**, *9*, doi:10.1093/aobpla/plx030.
20. 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*, 1–24, doi:10.1371/journal.pone.0180186.
21. Dayanne, D. Multi-study inference of regulatory networks for more accurate models of gene regulation. *PLoS Comput. Biol.* **2019**, *15*, 1–22, doi:10.1101/279224.
22. Needham, C.J.; Manfield, I.W.; Bulpitt, A.J.; Gilmartin, P.M.; Westhead, D.R. From gene expression to gene regulatory networks in Arabidopsis thaliana. *BMC Syst. Biol.* **2009**, *3*, 85, doi:10.1186/1752-0509-3-85.
23. Mochida, K.; Koda, S.; Inoue, K.; Nishii, R. Statistical and machine learning approaches to predict gene regulatory networks from transcriptome datasets. *Front. Plant Sci.* **2018**, *871*, 1–7, doi:10.3389/fpls.2018.01770.
24. Pratapa, A.; Jaliyal, A.P.; Law, J.N.; Bharadwaj, A.; Murali, T.M. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nat. Methods* **2019**, *17*, 147–154, doi:10.1101/642926.
25. De Lucas, M.; Brady, S.M. Gene regulatory networks in the Arabidopsis root. *Curr. Opin. Plant Biol.* **2013**, *16*, 50–55, doi:10.1016/j.pbi.2012.10.007.
26. El-Showk, S.; Taylor-Teeple, M.; Helariutta, Y.; Brady, S.M. Gene regulatory networks during Arabidopsis root vascular development. *Int. J. Plant Sci.* **2013**, *174*, 1090–1097, doi:10.1086/671449.
27. Taylor-Teeple, M.; Lin, L.; de Lucas, M.; Turco, G.; Toal, T.W.; Gaudinier, A.; Young, N.F.; Trabucco, G.M.; Veling, M.T.; Lamothe, R.; et al. An Arabidopsis gene regulatory network for secondary cell wall synthesis. *Nature* **2015**, *517*, 571–575, doi:10.1038/nature14099.
28. Doroshkov, A.V.; Konstantinov, D.K.; Afonnikov, D.A.; Gunbin, K.V. The evolution of gene regulatory networks controlling Arabidopsis thaliana L. trichome development. *BMC Plant Biol.* **2019**, *19*, 53, doi:10.1186/s12870-019-1640-2.
29. Hehl, R.; Bulow, L. *AthaMap Web Tools for the Analysis of Transcriptional and Posttranscriptional Regulation of Gene Expression in Arabidopsis thaliana*. *Plant Circadian Networks*; Methods in Molecular Biology; Humana Press: New York, NY, USA, 2014; Volume 1158.
30. Biggs, N. *Algebraic Graph Theory*, 2nd ed.; Cambridge University Press: Cambridge, UK, 1993.
31. Du, Q.; Campbell, M.; Yu, H.; Liu, K.; Walia, H.; Zhang, Q.; Zhang, C. Network-based feature selection reveals substructures of gene modules responding to salt stress in rice. *Plant Direct* **2019**, *3*, e00154, doi:10.1002/pld3.154.

32. 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, doi:10.1371/journal.pcbi.1006369.
33. Serre, J.-P. *Trees*; Springer: 1980.
34. Kleinberg, J.M. Authoritative sources in a hyperlinked environment. *J. ACM* **1999**, *46*, 604–632, doi:10.1145/324133.324140.
35. Lei, X.; Wang, S.; Wu, F.X. Identification of essential proteins based on improved HITS algorithm. *Genes* **2019**, *10*, 177, doi:10.3390/genes10020177.
36. Easley, D.; Kleinberg, J. *Networks, Crowds, and Markets*; Cambridge University Press: Cambridge, UK, 2010.
37. 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, 9–13 September 2001; pp. 444–445.
38. Langville, A.N.; Meyer, C.D. A survey of eigenvector methods for web information retrieval. *SIAM Rev.* **2005**, *47*, 135–161.
39. Emms, D.M.; Kelly, S. OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **2015**, *16*, 157, doi:10.1186/s13059-015-0721-2.
40. Bian, T.; Hu, J.; Deng, Y. Identifying influential nodes in complex networks based on AHP. *Phys. A Stat. Mech. Its Appl.* **2017**, *479*, 422–436, doi:10.1016/j.physa.2017.02.085.
41. Urbinati, A.; Galimberti, E.; Ruffo, G. Hubs and authorities of the scientific migration network. *arXiv* **2019**, arXiv:1907.07175.
42. Eldén, L. *Matrix Methods in Data Mining and Pattern Recognition*, 15th ed.; Siam: 2019.
43. Cowen, L.; Ideker, T.; Raphael, B.J.; Sharan, R. Network propagation: A universal amplifier of genetic associations. *Nat. Rev. Genet.* **2017**, *18*, 551.
44. Opsahl, T.; Agneessens, F.; Skvoretz, J. Node centrality in weighted networks: Generalizing degree and shortest paths. *Soc. Networks* **2010**, *32*, 245–251.
45. Gazestani, V.H.; Lewis, N.E. From genotype to phenotype: Augmenting deep learning with networks and systems biology. *Curr. Opin. Syst. Biol.* **2019**, *15*, 68–73.
46. Loscalzo, J.; Barabási, A.-L. *Network Science*, 1st ed.; Cambridge University Press: Cambridge, UK, 2016.
47. Janwa, H.; Massey, S.E.; Velez, J.; Mishra, B. On the origin of biomolecular networks. *Front. Genet.* **2019**, *10*, 240–257, doi:10.3389/fgene.2019.00240.
48. Aukerman, M.J.; Hirschfeld, M.; Wester, L.; Weaver, M.; Clack, T.; Amasino, R.M.; Sharrock, R.A. A deletion in the PHYD gene of the arabidopsis wassilewskija ecotype defines a role for phytochrome D in red/far-red light sensing. *Plant Cell* **1997**, *9*, 1317–1326, doi:10.1105/tpc.9.8.1317.