Research Article

Research on Transboundary Regulation of Plant-Derived Exogenous MiRNA Based on Biological Big Data

Zhi Li^[1], Xu Wei,² Shuyi Li,³ Jiashi Zhao,¹ Xiang Li,¹ and Liwan Zhu¹

¹School of Computer Science and Technology, Changchun University of Science and Technology, Changchun 130022, China ²College of Computer Science and Technology, Jilin University, Changchun 130012, China ³School of Foreign Languages, Harbin Institute of Technology, Harbin 150001, China

Correspondence should be addressed to Zhi Li; 153659073@qq.com

Received 27 November 2020; Revised 6 January 2021; Accepted 13 January 2021; Published 31 January 2021

Academic Editor: Le Zhang

Copyright © 2021 Zhi Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In recent years, researchers have discovered plant miRNA (plant xenomiR) in mammalian samples, but it is unclear whether it exists stably and participates in regulation. In this paper, a cross-border regulation model of plant miRNAs based on biological big data is constructed to study the possible cross-border regulation of plant miRNAs. Firstly, a variety of human edible plants were selected, and based on the miRNA data detected in human experimental studies, screening was performed to obtain the plant xenomiR that may stably exist in the human body. Then, we use plant and animal target gene prediction methods to obtain the mRNAs of animals and plants that may be regulated, respectively. Finally, we use GO (Gene Ontology) and the Multiple Dimensional Scaling (MDS) algorithm to analyze the biological processes regulated by plants and animals. We obtain the relationship between different biological processes and explore the regulatory commonality and individuality of plant xenomiR in plants. Studies have shown that the development and metabolic functions of the human body are affected by daily eating habits. Soybeans, corn, and rice can not only affect the daily development and metabolism of the human body but also regulate biological processes such as protein modification and mitosis. This conclusion explains the reasons for the different physiological functions of the human body. This research is an important meaning for the design of small RNA drugs in Chinese herbal medicine and the treatment of human nutritional diseases.

1. Introduction

MicroRNA (miRNA) is a type of noncoding single-stranded small RNA with a size of 21–23 nucleotides. It binds to mRNA through the base complementation rule, degrades mRNA, inhibits mRNA translation, and ultimately regulates gene expression [1]. It plays a role in all aspects of the life cycle of biological cells. With the development of genome sequencing technology and the emergence of miRNA verification methods based on high-throughput biological big data, researchers have discovered more miRNAs and studied them in depth. The expression regulation function of miRNA has become a research hotspot in this field [2].

Studies have shown that miRNAs can not only function in their own body but also perform cross-border regulation. As an exogenous biological activity unit, it can affect the expression of heterologous mRNA through base complementation. In the process of studying the regulatory functions of miRNAs, plant miRNAs were found in mammalian samples [3–8]. Researchers infer that these plant miRNAs enter the animal's body through food. Researchers call these exogenous miRNAs of plant origin (plant xenomiR).

Although there are many studies on plant xenomiR transboundary regulation, they are all based on a single plant miRNA. However, studies have found that not all plant miRNAs can enter the human body. The human body specifically absorbs plant miRNA. Therefore, it is meaningful to select plant miRNAs that exist in animals for subsequent analysis. The determination of plant xenomiR is the key to this research [9–11].

In this article, on the basis of plant xenomiR transboundary regulation, a research model of plant xenomiR transboundary regulation based on biological big data is proposed and constructed. The possible regulatory functions of plant XenomiR in animals and plants were studied, respectively, and the Multiple Dimensional Scaling (MDS) algorithm was used to analyze the regulatory functions of different species, to study the cross-species regulation mechanism of XenomiR on different kinds of edible plants. Not only the regulation effect of plant xenomiR on the plant body but also its regulation effect on the human body can be obtained. Explore these plant xenomiR regulatory characteristics. This model can be applied to the research of crossspecies regulation mechanism, Chinese herbal medicine efficacy, nutrition, and other fields.

2. Materials and Methods

2.1. Technical Route. In this paper, a research model of plant xenomiR transboundary regulation based on biological big data is constructed. Firstly, select the miRNA data of various human edible plants, and use the miRNA data detected in the human body obtained by the second-generation sequencing technology to perform data screening to obtain the plant xenomiR data present in human samples. Then, using plant and animal target gene prediction methods, the plant and animal mRNA that may be regulated are obtained, respectively. Finally, perform functional analysis of mRNA, use GO enrichment analysis and the Multiple Dimensional Scaling (MDS) algorithm to analyze its biological processes in plants and animals from the same perspective, and obtain the relationship between different biological processes. Look for the regulatory commonalities and individualities of plant miRNAs that coexist in plants and humans.

The technical route is shown in Figure 1.

- (1) Data acquisition: choose a variety of plants that are edible by humans. Obtain human miRNA secondgeneration sequencing data, miRNA sequence data, and mRNA data of related plants from professional public databases and literature data. Search for data related to plant miRNA in human miRNA secondgeneration sequencing data.
- (2) Target gene prediction: use the target gene prediction algorithm for xenomiR mRNA comparison and statistical screening to obtain the target gene data of plant xenomiR in plants and humans.
- (3) Core node screening: use the LeaderRank algorithm to calculate the score of each node and find the core node of the network.
- (4) Core network construction and function enrichment analysis: build a biological regulatory network through core nodes and combine it with function enrichment analysis to study the similarities and differences between these miRNAs participating in the biological processes of plants and humans.
- (5) GO analysis: analyze the similarities and differences of the data in different comparison groups, and analyze the GO semantic relationship of these data by using the MDS algorithm. Explore the commonalities and individual differences of its function of gene regulation.

2.2. Data Acquisition and Preprocessing. Data needed for study: miRNA and mRNA of edible plants (soybean, rice, and corn), sequencing data based on high-throughput human miRNA, human mRNA data, gene annotation files of edible plants and humans.

The miRNA data of plant crops comes from the miRBase database (http://www.mirbase.org/); the database version is 22.1 (October 2018). The study is to explore the role of plant miRNAs obtained by humans through food. Therefore, the necessary conditions for selecting crops for research are as follows: crops that are often eaten by humans in daily life and related data are perfect for subsequent analysis. After screening, the final eligible crops are as follows: soybeans, rice, and corn.

Related information is shown in Table 1.

The research needs to obtain plant miRNA data found in human samples. Qi Zhao and others in our laboratory analyzed 388 human small RNA sequencing data and detected a total of 484 plant miRNAs, including 166 unique miRNA sequences [12].

The relevant plant mRNA data were downloaded from NCBI (https://www.ncbi.nlm.nih.gov/), and the human mRNA data were downloaded from GENCODE (https:// ucscgenomics.soe.ucsc.edu/gencode/). The relevant data statistics are shown in Table 2.

Compare the obtained miRNA data of soybean, rice, and corn with the 166 plant miRNA data found in human samples to obtain the source of plant xenomiR. The statistical results are shown in Table 3.

2.3. Target Gene Prediction. Find the target genes of plants and humans corresponding to the target of the corresponding plant xenomiR.

In this paper, the tapir algorithm based on RNAHybrid was used to predict the target gene [13, 14]. The Tapir algorithm can set its own parameters to make the results more accurate. The Tapir algorithm is specially designed for plant miRNAs and has a good predictive effect on plant miRNA target genes. The screening criteria used are as follows: exact match of seed region; remove the results with more than 3 bases of mismatches; set the MFE (minimum free energy) threshold of the result to -25; and set the *p_value* threshold to 0.05. When the above conditions are met, it is deemed to meet the standards.

The prediction process used in this paper is shown in Figure 2.

The relevant statistical results are shown in Table 4.

Since three edible plants were selected in this article, all experimental results in this article are divided into six groups.

2.4. Core Node Screening. The currently acquired data contains a large number of independent nodes or nodes that are not closely related to other network nodes. This information makes the biological pathways unobvious and fails to reflect the core of the network, making it difficult to understand the regulatory network. Therefore, it is necessary to find the core node after filtering the nodes.



FIGURE 1: Research model of plant xenomiR transboundary regulation based on biological big data.

TABLE 1: Plant information table.

Name	Scientific name	Quantity
Soybeans	Glycine max (Linn.) Merr.	756
Rice	Oryza sativa japonica group	738
corn	Zea mays Linn.	325

TABLE 2: Plant, crop, and human mRNA data.

Species name	Soybean	Rice	Corn	Human
Quantity	71149	42474	56152	206694

TABLE 3: Statistics of plant xenomiR.

Plant name	Soybean	Rice	Corn
Quantity	42	46	29
Percentage	5.56%	6.23%	8.92%



FIGURE 2: Plant xenomiR target gene prediction process.

The LeaderRank algorithm is used to score and rank the nodes required by the biological regulatory network. It adds the background node b based on the PageRank algorithm and connects it with other nodes to form a strong connection graph. The degree of each node is greater than 0, to avoid the existence of isolated nodes, so that the results can converge faster [15, 16].

Set the LeaderRank score of node *b* as $S_b = 0$ and the scores of other *n* nodes as $S_i = 1$, through the iterative formula:

$$S_{i}(t) = \sum_{j=1}^{n+1} \frac{a_{ji}}{O_{j}} S_{j}(t-1).$$
(1)

After the element value gradually converges, we use t_c to represent the number of convergence and then obtain the score of each node:

$$S_i = S_i(t_c) + \frac{S_b(t_c)}{n}.$$
 (2)

After using the above method to operate, the changes in the number of nodes before and after are shown in Table 5.

2.5. GO Analysis Based on Multiple Dimensional Scaling Algorithm. By performing functional enrichment analysis on multiple sets of data, information such as the contained biological processes can be obtained. However, usually due to the diversity of results, it is difficult to dig out meaningful biological information from a large amount of information.

The dimensionality reduction processing of the data through the MDS algorithm can ensure that the distance between the data samples after dimensionality reduction does not change compared with that before dimensionality reduction, which is very helpful for GO semantic analysis [17].

In the MDS algorithm, assuming that there are m samples in the data, their sample space is

TABLE 4: Statistics of target gene prediction results.

	Soybean-soybean	Rice-rice	Corn-corn	Soybean-person	Rice-person	Corn-person
Target	1119	767	443	1419	2409	1343
GeneID	380	396	214	378	632	358

TABLE 5: Statistics of the number of nodes before and after screening.

	Soybean-soybean	Rice-rice	Corn-corn	Soybean-person	Rice-person	Corn-person
Before screening	380	396	214	378	632	358
After screening	156	129	54	260	231	43

$$T = \{x_1, x_2, x_3, x_4, \dots, x_m\} x_i \in \mathbb{R}^d.$$
(3)

Assuming that the matrix D is the distance between samples and $D \in \mathbb{R}^{m \times m}$, the distance between sample x_i and sample x_j is the element d_{ij} in D. The sample in the new space can be expressed as

$$Z \in \mathbb{R}^{d' \times m},$$

$$d' \le d.$$
(4)

The Euclidean distance between the new and old data in the two spaces is the same as

$$Z_i - Z_j = d_{ij}.$$
 (5)

Let the inner product matrix of the sample after dimensionality reduction be

$$B = Z^T Z \in R^{m \times m}.$$
 (6)

And
$$b_{ij} = Z_i^T Z_j$$
; then,
 $d_{ij}^2 = Z_i^2 + Z_j^2 - 2Z_i^T Z_j = b_{ii} + b_{jj} - 2b_{ij}.$ (7)

When the sample Z after dimensionality reduction is centered $\sum_{i=1}^{m} Z_i = 0$, the sum of the rows and columns of matrix B is 0; that is,

$$\sum_{i=1}^{m} b_{ij} = \sum_{j=1}^{m} b_{ij} = 0.$$
 (8)

When using tr () to represent the matrix,

$$trB = \sum_{j=1}^{m} Z_i^2.$$
 (9)

Available:

$$\sum_{i=1}^{m} d_{ij}^{2} = tr(B) + mb_{jj},$$
(10)

$$\sum_{j=1}^{m} d_{ij}^2 = tr(B) + mb_{ii},$$
(11)

$$\sum_{i=1}^{m} \sum_{j=1}^{m} d_{ij}^{2} = 2mtr(B).$$
(12)

$$d_{i.}^{2} = \frac{1}{m} \sum_{j=1}^{m} d_{ij}^{2},$$

$$d_{.j}^{2} = \frac{1}{m} \sum_{i=1}^{m} d_{ij}^{2},$$
(13)

$$d_{..}^{2} = \frac{1}{m^{2}} \sum_{i=1}^{m} \sum_{j=1}^{m} d_{ij}^{2}.$$
 (14)

In summary,

$$b_{ij} = -\frac{1}{2} \Big(d_{ij}^2 - d_{i.}^2 - d_{.j}^2 + d_{..}^2 \Big).$$
(15)

In this algorithm, D is kept unchanged, that is, the distance between samples is unchanged, and so the inner product matrix B can be calculated by D. At the same time, because B is a symmetric matrix, it can be feature decomposition, and we can get

$$B = V \Delta V_T, \tag{16}$$

where \wedge is the diagonal matrix, which is composed of eigenvalues, and V is the corresponding eigenvector matrix.

Assuming that there are d_* nonzero eigenvalues, they can form a diagonal matrix Λ_* . Assuming Λ_* is the corresponding eigenvector matrix, the sample in the new space represents Z:

$$Z = \Lambda_*^{1/2} V_*^T \in \mathbb{R}^{d^* \times m}.$$
(17)

3. Results and Discussion

3.1. Functional Enrichment Analysis of Plant-Plant Group. For the plant's own regulation mode, we select the data with $p_value \le 0.05$ for analysis, take $\log_{10} p_value$, and sort them from large to small, combined with the number of genes in the relevant biological process.

The statistical results of soybean-soybean biological process information are shown in Figure 3.

The statistical results of the soybean protein interaction network are shown in Figure 4.

The statistical results of rice-rice biological process information are shown in Figure 5.

The statistical results of the rice-protein interaction network are shown in Figure 6.

Make



FIGURE 3: Soybean-soybean biological process information statistics map (total of 33 items).



FIGURE 4: Soybean protein interaction network.



FIGURE 5: Rice-Rice biological process information statistics map (total of 59 items).



FIGURE 6: Rice-protein interaction network (metabolism-related).

The statistical results of corn-corn biological process information are shown in Figure 7.

The statistical results of the corn protein interaction network are shown in Figure 8.

3.2. Functional Enrichment Analysis of Plant-Human Group. In the experiment of plant and colleagues, the data $p_$ value ≤ 0.05 are also selected for analysis.

The statistical results of soybean-human biological process information are shown in Figure 9.

The statistical results of the soybean-human protein interaction network are shown in Figure 10.

According to its biological process information, rice has a particularly obvious regulation in human development. The protein interaction network for nervous system development is shown in Figure 11.

The statistical results of rice-human biological process information are shown in Figure 11.

The statistical results of rice-human-protein interaction network are shown in Figure 12.

The statistical results of corn-human biological process information are shown in Figure 13.

The statistical results of corn-human-protein interaction network are shown in Figure 14.

3.3. Biological Process and GO Semantic Analysis. In the GO semantic analysis graph, color is used to represent the change of $\log_{10} p$ -value, the larger the value, the warmer the color, and the radius is used to represent the value of $\log size$.

3.3.1. Soybean. Soybean miRNAs play different regulatory roles in plants and humans, and their regulatory roles in humans are quite different.

Soybean miRAN plays a regulatory role in soybean that is mostly related to plant development, such as leaf development, integument development, another development, and positive developmental regulation (see Figure 15).

The regulatory role of soybean miRNA in the human body is related to cell protein modification process, cell response to external stimuli, cell protein metabolism process, protein ubiquitination, regulation of mitotic cell cycle, and so forth. There is no obvious correlation between various biological processes (see Figure 16).

3.3.2. Rice. Rice miRNAs regulate the biological metabolism and development of rice itself, as well as humans. At the same time, it can also regulate the response of cells to external stimuli, cell processes, and cell communication.

Most of the regulatory effects of rice miRNA are related to metabolism and development. Metabolism includes the metabolic process of cellular aromatic compounds, the metabolic process of organic cyclic compounds, the cellular metabolic process, the cellular macromolecular metabolic process, and the organic matter metabolic processes. Development-related includes multicellular organism development, flower development, postembryonic development, system development, organ development, reproductive system development, postembryonic organ development, and tissue development (see Figure 17)

Most of the regulatory effects of rice miRNA in the human body are also related to metabolism and development. Metabolism includes organic matter metabolism, primary metabolism, cellular macromolecular metabolism, cell metabolism, nitrogen compound metabolism, and phosphorus metabolism process. Development-related includes nervous system development, neuron production, cell differentiation, brain development and ventricular system development, and lateral ventricle development. At the same time, it can also regulate the response of cells to stimuli and regulate cell processes, cell-to-cell communication, and cell division cycles (see Figure 18).

3.3.3. Corn. Maize miRNAs can regulate metabolism in both plants and humans, and they can also regulate biological processes.

Most of the regulatory effects of maize miRNA in maize are related to metabolism and transcription and translation, such as lignin catabolism, phenyl propane catabolism, cellular aromatic compound metabolism, phenyl propane metabolism, and organic cyclic compound metabolism. Transcription and translation-related include nucleic acid template transcription, biosynthetic process, nucleic acid template transcription, regulation of gene expression in biosynthesis process, regulation of nucleic acid template transcription, and regulation of RNA biosynthesis process (see Figure 19).

The miRNA of corn can regulate the response to stimuli in the human body, such as cell response to endogenous stimuli, cell response to peptide hormone stimulation, cell response to growth factor stimulation, cell response to organic matter, and cell response reaction to chemical stimuli. It can regulate metabolism, such as



FIGURE 7: Corn-corn biological process information statistics map (total of 37 entries).



FIGURE 8: Corn protein interaction network.



FIGURE 9: Soybean-human biological process information statistics map (total of 7 entries).



FIGURE 10: Soybean-human protein interaction network (related to protein ubiquitination).



FIGURE 11: Rice-human biological process information statistics map (total of 90 entries showing only partial information).



FIGURE 12: Rice-human-protein interaction network (related to nervous system development).



FIGURE 13: Corn-human biological process information statistics map (total of 86 entries showing only partial information).



FIGURE 14: Corn-human-protein interaction network.



FIGURE 15: Soybean-soybean GO semantic analysis diagram.



FIGURE 16: Semantic analysis diagram of soybean-human GO.



FIGURE 17: Semantic analyses of rice GO.

cell protein metabolism process, positive regulation of phosphoric acid metabolism process, and organic nitrogen compound metabolism process. It can regulate cell functions, such as regulation of multicellular tissue processes and regulation of cytokine production (see Figure 20).



FIGURE 18: Semantic analysis diagram of rice-human GO.



FIGURE 19: Semantic analysis diagram of corn-maize GO.



FIGURE 20: Semantic analysis diagram of corn-human GO.

4. Conclusions

By research, it is found that, in terms of plant xenomiR's, common regulation of humans, soybean, rice, and corn all, contain miRNAs that can regulate the daily development and metabolism of the human body. These miRNAs can also regulate other biological processes, such as response to external stimuli (GO: 0051716), protein modification (GO: 0006464), and mitotic cell cycle (GO: 0007346).

In terms of personality regulation, soybean can regulate the process of protein ubiquitination (GO: 0016567), which plays an important role in protein metabolism and degradation and also participates in cell proliferation and differentiation, repairing damage, and immune inflammation. Rice is particularly effective in regulating human development, such as nervous system development (GO: 0007399), brain development (GO: 0007420), and ventricular system development (GO: 0021591). Corn has a regulatory effect on protein phosphorylation (GO: 0001932), which can regulate the activity and function of the protein, which is the most common and important regulatory mechanism.

This shows that when humans eat plants, the development and metabolic functions of the human body will be affected accordingly, and various physiological functions will also change. And due to different eating habits, the possible adjustments will also be different.

This conclusion explains the different effects of daily eating habits on the physiological functions of the human body, which is consistent with the experimental research results of Sanchta [18]. Plant xenomiR can play a role in the human body, and edible plant xenomiR can be used as a nutritional supplement, which can bring beneficial effects to human health.

Journal of Healthcare Engineering

Data Availability

All personnel can access the data that include plant miRNA and mRNA information and related processing procedures. The data download address is https://pan.baidu.com/s/ 1YIOzS4LsWEmQNuTln7yemA with code '5b2e'.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This study was financed by the Key Research and Development Projects of China's Jilin Province Science and Technology Development Plan (no. 20200401078GX).

References

- V. N. Kim and V. Narry, "MicroRNA biogenesis: coordinated cropping and dicing," *Nature Reviews Molecular Cell Biology*, vol. 6, no. 5, pp. 376–385, 2005.
- [2] S. Altuvia, "Identification of bacterial small non-coding RNAs: experimental approaches," *Current Opinion in Microbiology*, vol. 10, no. 3, pp. 257–261, 2007.
- [3] C. De la Rosa and A. C. Reyes, "Origin and evolutionary dynamics of the miR2119 and ADH1 regulatory module in legumes," *Genome Biology and Evolution*, vol. 12, no. 12, pp. 2355–2369, 2020.
- [4] E. Layton, "Anna marie fairhurst. Sam griffiths jones. Richard K grencis. Ian S roberts, "regulatory RNAs: a universal language for inter-domain communication," *International Journal of Molecular Sciences*, vol. 21, no. 23, 2020.
- [5] Q. Cai, L. Qiao, M. Wang et al., "Plants send small RNAs in extracellular vesicles to fungal pathogen to silence virulence genes," *Science*, vol. 360, no. 6393, pp. 1126–1129, 2018.
- [6] A. Aucher, D. M. Rudnicka, and D. M. Davis, "MicroRNAs transfer from human macrophages to hepato-carcinoma cells and inhibit proliferation," *The Journal of Immunology*, vol. 191, no. 12, pp. 6250–6260, 2013.
- [7] M. Zhan and H. Chang, "Research progress of exogenous plant mirnas in crosskingdom regulation," *Current Bioinformatics*, vol. 14, no. 3, pp. 214–245, 2019.
- [8] L. Zhang, D. Hou, X. Chen et al., "Exogenous plant MIR168a specifically targets mammalian LDLRAP1: evidence of crosskingdom regulation by microRNA," *Cell Research*, vol. 22, no. 1, pp. 107–126, 2012.
- [9] X. Chen, H. Liang, J. Zhang, K. Zen, and C.-Y. Zhang, "Secreted microRNAs: a new form of intercellular communication," *Trends in Cell Biology*, vol. 22, no. 3, pp. 125–132, 2012.
- [10] J. Shu, K. Chiang, and D. Zhao, "Human absorbable microrna prediction based on an ensemble manifold ranking model," in *Proceedings of the Bioinformatics and Biomedicine (BIBM)*, 2015 IEEE International Conference on, IEEE, pp. 295–300, Hong Kong, China, June 2015.
- [11] J. Shu, K. Chiang, J. Zempleni et al., "Computational characterization of exogenous microRNAs that can be transferred into human circulation," *PLoS ONE*, vol. 10, no. 11, Article ID e0140587, 2015.

- [12] Z. Qi, L. Yuanning, Z. Ning et al., "Evidence for plant-derived xenomiRs based on a largescale analysis of public small RNA sequencing data from human samples," *PLoS ONE*, vol. 13, no. 6, Article ID e0187519, 2018.
- [13] B. P. Lewis, C. B. Burge, and D. P. Bartel, "Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets," *Cell*, vol. 120, no. 1, pp. 15–20, 2005.
- [14] M. Rehmsmeier, P. Steffen, M. Höchsmann et al., "Fast and effective prediction of microRNA/target duplexes," *RNA*, vol. 10, no. 10, pp. 1507–1517, 2004.
- [15] S. Brin and L. Page, "The anatomy of a large-scale hypertextual Web search engine," *Computer Networks and ISDN Systems*, vol. 30, no. 1–7, pp. 107–117, 1998.
- [16] L. Ü. Linyuan, Y. C. Zhang, C. H. Yeung et al., "Leaders in social networks the delicious case," *PLoS ONE*, vol. 6, no. 6, Article ID e21202, 2011.
- [17] X. Liu, Y. Hu, S. North, Y. Hu, S. North, and H.-W. Shen, "CorrelatedMultiples: spatially coherent small multiples with constrained multi-dimensional scaling," *Computer Graphics Forum*, vol. 37, no. 1, pp. 7–18, 2018.
- [18] R. A. M. H. Sanchta, "Dietary plant miRNAs as an augmented therapy: cross-kingdom gene regulation," *RNA Biology*, vol. 15, no. 12, pp. 1433–1439, 2018.