ABSTRACT
of the dissertation for the degree
Doctor of Philosophy (PhD)
6D070100 - Biotechnology
Rakhmetullina Aizhan Kazievna

Characteristics of miRNAs binding with mRNAs of transcription factor genes of agricultural plants

General description of the research.

This thesis is devoted to the determination of the quantitative characteristics of the interaction of miRNAs with mRNAs of genes of transcription factors of agricultural plants.

Significance of the research.

Transcription factors (TF) play a key role in the regulation of the expression of many genes and, in general, the entire eukaryotic genome. Currently, there is an important problem of controlling the expression of TF in order to increase the productivity of plants by changing their growth, development, and increasing resistance to biotic and abiotic environmental factors. Most types of agricultural plants contain more than 50 families of TF and it is necessary to choose the most important ones in the functioning of plants.

According to the Food and Agriculture Organization of the United Nations (FAO), the most used grain crops in the world are rice, corn and wheat, so these types of plants are chosen as objects of our research. We used Arabidopsis as the most well-studied plant type.

Recently, miRNA (mRNA-inhibitory RNA) effect on TF expression has been discovered, which opens up new possibilities for regulating their synthesis. The miRNAs can interact with mRNA of TF genes with different efficiencies, defining a wide range from weak suppression of expression to complete repression of protein synthesis or mRNA degradation. Solving the problem of establishing the interaction of miRNA with mRNA target genes is complicated by the determination of miRNA target genes and the establishment of quantitative characteristics of the interaction of these molecules. Existing programs for predicting target genes and miRNA binding sites in their mRNAs are imperfect and find many false-positive target genes and miRNA sites. This reason significantly inhibits the establishment of miRNA associations and their target genes in plants and animals.

Despite the constant increase in the detected miRNAs in plant genomes, it is not always possible to credibly identify their specific target genes. Data on changes in the concentration of miRNA in cells at different stages of ontogenesis and under various influences are rapidly increasing, which confirms their important role in plant functioning. The use of miRNAs as modifiers of gene expression regulation is promising, however, it is necessary to accurately identify the target genes and avoid side effects. The MirTarget program used in the work can solve this problem

and therefore the goal and research tasks set in the dissertation can be solved. The establishment of miRNA associations and their target genes will enhance the efficiency of plant genetic engineering.

The purpose of the research: To study the quantitative characteristics of the interaction of miRNA with mRNA of TF genes of *A. thaliana, O. sativa, Z. mays* and *T. aestivum* and to identify the most effective associations of miRNA and target genes.

The main tasks of the research:

1. To create databases of TF genes of the TCP, HSF, MYB, GRAS, ERF, C2H2 families and miRNA of *A. thaliana, O. sativa, Z. mays* and *T. aestivum*.
2. To determine the characteristics of interaction miRNA with mRNA of the TCP TF genes of *A. thaliana, O. sativa, Z. mays* and *T. aestivum*.
3. To determine the characteristics of interaction miRNA with mRNA of the HSF TF genes of *A. thaliana, O. sativa, Z. mays* and *T. aestivum*.
4. To determine the characteristics of interaction miRNA with mRNA of the MYB TF genes of *A. thaliana, O. sativa, Z. mays* and *T. aestivum*.
5. To determine the characteristics of interaction miRNA with mRNA of the GRAS TF genes of *A. thaliana, O. sativa, Z. mays* and *T. aestivum*.
6. To determine the characteristics of interaction miRNA with mRNA of the ERF TF genes of *A. thaliana, O. sativa, Z. mays* and *T. aestivum*.
7. To determine the characteristics of interaction miRNA with mRNA of the C2H2 TF genes of *A. thaliana, O. sativa, Z. mays* and *T. aestivum*.
8. To determine characteristics of interaction rice, maize and wheat miRNAs with the mRNA of human genes.

Objects of the research: miRNA and TF genes of fully sequenced plant genomes of *A. thaliana, O. sativa, Z. mays* and *T. aestivum;* human genes as plant miRNA targets.

Subject of the study: Quantitative characteristics of the interaction of miRNA with mRNA of TF genes of *A. thaliana, O. sativa, Z. mays* and *T. aestivum*

Methods of the study: Methods of computer analysis and search for miRNA binding sites in mRNAs of genes based on modeling hydrogen bonds using the MiRTarget program.

The scientific novelty of the research.

The novelty of the study is the following: 1. The groups of target genes for miRNAs among TF families TCP, HSF, MYB, GRAS, ERF, C2H2 of the *A. thaliana, O. sativa, Z. mays* and *T. aestivum* have been predicted for the first time; 2. For the first time binding sites of some miRNAs were identified in mRNAs of the target genes of different plants, that are conservative and encode conservative oligopeptides; 3. For the first time, it was revealed that some miRNAs can have more than one target gene and mRNAs of some target genes contain binding sites for two or more different miRNAs; 4. The binding sites of some miRNAs in CDS mRNA of plant TF genes can encode different oligopeptides depending on the reading frame; 5. For the first time, the binding sites of miRNAs of rice, maize

and wheat with mRNA genes involved in various human diseases have been established.

The theoretical significance of the research. Quantitative characteristics of the interaction of miRNA with mRNA of many plant TF genes have been established, which make it possible to evaluate the interactions of these molecules. Fully complementary interactions of nucleotides of some miRNAs with mRNA of plant TF genes were revealed, which indicates the efficient binding of these miRNAs to mRNAs. The revealed interactions of plant miRNA with TF mRNA contribute to understanding the effect of miRNA on many physiological processes in plants and make it possible to purposefully use miRNA in genetic modification of plants. The predicted interactions of plant miRNAs with human mRNA genes make it possible to recommend some miRNAs as regulators of human gene expression.

The practical significance of the research. The establishment of plant miRNA associations and their target genes TFs of the TCP, MYB, GRAS, ERF, C2H2 families reduces the time and material costs of searching for connections between miRNAs and target genes of agricultural plants, compared to finding associations of miRNAs with mRNAs experimentally. The revealed associations of miRNA and TF target genes in the studied plants allow targeted modification to improve many useful properties of various species of plants. MirTarget program has been successfully used to find miRNA binding sites in mRNAs of plant and human genes, and to quantify the interaction of miRNA with mRNA target genes. The identified plant miRNAs can influence human genes and be used in medicine as biologically active compounds, since they easily penetrate the human body. The copyright certificate “MirTarSeq” No. 15600 dated March 2, 2021 was received.

The main provisions for the defence:

Some genes out of the total number of 2403 genes of TCP, HSF, MYB, GRAS, ERF, C2H2 TF of *A. thaliana, O. sativa, Z. mays* and *T. aestivum* families can be miRNA targets of these organisms.

The results of *in silico* studies of the influence of plant miRNAs on mRNA of TF genes of TCP, HSF, MYB, GRAS, ERF, C2H2 families of *A. thaliana, O. sativa, Z. mays* and *T. aestivum* showed that the many studied miRNAs can influence plant growth and development.

The miRNA binding sites in the mRNA of the studied genes of TCP, HSF, MYB, GRAS, ERF, C2H2 TF are located in the 5'UTR, CDS, and 3'UTR.

Different quantitative characteristics of the interaction between miRNA and mRNA of the studied TF genes of the TCP, HSF, MYB, GRAS, ERF, C2H2 families indicate the effectiveness of the influence of miRNA on a number of genes of these TF families.

Most of the studied miRNAs of *O. sativa, T. aestivum, Z. mays,* which are actively involved in plant growth and development, can affect mRNA translation of human genes.

The main research results and conclusions:

1. In the format available for the MirTarget program, the following databases were created: for genes of the TCP, HSF, MYB, GRAS, ERF, C2H2 TF families,

consisting of 442 genes of *A. thaliana*, 474 genes of *O. sativa*, 653 genes of *Z. mays* and 834 genes of *T. aestivum*; for 428, 738, 325, and 125 miRNAs of *A. thaliana, O. sativa, Z. mays* and *T. aestivum*, respectively.

2. Based on the quantitative characteristics of the interaction of 428 ath-miRNAs and mRNA of 27 TCP family genes of *A. thaliana* revealed only 11 target genes for five miRNAs. The ath-miR5021-5p binding sites were found in mRNAs of three genes of *A. thaliana* and in mRNAs of 27 genes of 17 plant species. The ath-miR5658-5p binding sites in mRNA of 23 genes of 19 plant species are conservative. Of 738 osa-miRNAs and mRNA of 22 TCP family genes of *O. sativa* only 14 genes were shown to be the targets for 17 miRNAs. The osa-miR2102-5p binding sites encode the AAAAAA and GGGGGG oligopeptides in mRNAs of ten genes of seven plant species and in mRNAs of eight genes of six plant species, respectively. The analysis of binding of 325 zma-miRNAs and mRNA of 46 TCP family genes of *Z. mays* predicted only seven target genes for 11 miRNAs. Of 125 miRNA and 28 mRNA of the TCP family genes of *T. aestivum* revealed only five tae-miRNAs have five mRNA target genes. The tae-miR319-3p binding sites encode the conservative oligopeptide QRGPLQS in the TCP TF in 54 plant species. The tae-miR444a-3p binding sites encode the STSETS oligopeptide in mRNAs of 29 genes of 28 plant species.

3. Of 428 ath-miRNAs and mRNA of 24 HSF genes of *A. thaliana*, only five genes were targets for five miRNAs. Of 738 osa-miRNAs and mRNA of 25 HSF family genes of *O. sativa* 12 genes were under the control of ten miRNAs, of which the largest number of binding sites had miR5075-3p. Of 325 zma-miRNAs and mRNA 28 HSF family genes of *Z. mays* it was revealed only six target genes for seven miRNAs. Of 125 tae-miRNAs and the mRNA of 51 genes of the *T. aestivum* HSF family, it was found that only three genes of the HSF family were targets for four miRNAs.

4. Out of 428 ath-miRNA binding sites in the mRNAs of 144 MYB genes of *A. thaliana* it was revealed that 32 genes are targets of 15 miRNAs. Of 738 osa-miRNA and mRNA of 124 genes of the MYB *O. sativa* family revealed that 34 genes were targets for 32 miRNAs. Of 325 zma-miRNAs and mRNA of 169 genes of the *Z. mays* MYB family, 25 genes were the targets for 26 miRNAs. Of 125 tae-miRNAs binding and mRNA of 258 MYB family genes of *T. aestivum* revealed that only eight genes were targets for eight miRNAs. The tae-miR159a,b-3p binding sites encode the WSSIRSK oligopeptide, which is conserved in the 27 proteins of the MYB TF for 22 plant species. The proteins of the MYB family of 22 plant species contained the ELPSNQ oligopeptide encoded by miR159e-3p binding sites in mRNAs of 23 genes of 20 plant species.

5. Of 428 ath-miRNAs and mRNA of 37 genes of the *A. thaliana* GRAS family, it was found that only 11 genes were targets for eight miRNAs. Of 738 osa-miRNAs, only 16 miRNAs could bind to mRNA of 18 genes from 60 genes of the GRAS *O. sativa* family. Of 325 zma-miRNAs and mRNA of 86 genes of the GRAS family of *Z. mays*, only 14 genes were targets for eight miRNAs. Of 125 tae-miRNAs interaction with mRNA of 117 GRAS TF genes of *T. aestivum* revealed only five target genes for three miRNAs. The nucleotide sequences ath-

miR171a-3p, osa-miR171a-3p, zma-miR171n-3p were identical and had binding sites in mRNA for 13 genes of the TF GRAS family and encode conservative ILARN oligopeptide.

6. Of 428 ath-miRNAs and mRNA of 123 genes of the ERF TF family of *A. thaliana*, it was revealed that 25 genes were targets for eight miRNAs. Of 738 *O. sativa* miRNAs, only 13 miRNAs effectively bound to mRNA of 16 genes from 138 genes. Of 325 miRNAs and mRNA of 186 genes of the ERF TF family of *Z. mays*, only two genes were targets for two miRNAs. Of 125 miRNAs and 169 mRNA of *T. aestivum* genes, it was found that only five genes were targets for four miRNAs.

7. Of 428 ath-miRNAs and mRNA of 87 C2H2 TF genes of *A. thaliana*, only 17 genes were targets of nine miRNAs. Of 738 osa-miRNAs and mRNAs of 105 C2H2 family genes of *O. sativa* revealed only 17 target genes for 14 miRNAs. Of the 325 zma-miRNAs and mRNA of 138 of the C2H2 TF family genes of *Z. mays*, the number of target genes that bind to zma-miRNAs was equal six. Of the 125 miRNAs and mRNA of 211 C2H2 family genes of *T. aestivum*, only six target genes were identified for three miRNAs.

8. For ath-miR5021-5p, ath-miR5658-5p, osa-miR2102-5p, osa-miR5075-3p, which have several target genes in the C2H2, ERF, GRAS families, it was found that the nucleotide sequences of the binding sites were conserved and encoded conservative oligopeptides: ath-miR5021-5p – SSSSSS, ath-miR5658-5p – HHHHH, osa-miR2102-5p – AAAAAA and GGGGGG, osa-miR5075-3p – AAAAA and GGGGGG.

9. Study of 325 zma-miRNAs interaction to mRNAs of 17508 human genes revealed only 38 target genes for nine single zma-miRNAs and 211 target genes for 94 zma-miRNA families. The study of 125 tae-miRNAs interaction with mRNA of 17508 human genes revealed only 116 target genes for 44 single tae-miRNAs and 57 target genes for 23 tae-miRNA families. Among 17,508 human genes with 738 osa-miRNAs, 942 target genes for 277 osa-miRNAs were established. A total of 641 target genes were identified for 131 single miRNAs. Most of the studied target genes of zma-miRNA, tae-miRNA, osa-miRNA may influence on human genes which participate in the development of oncological, neurodegenerative, and cardiovascular diseases.

The relationship of the research study with the scientific project.

The dissertation work was carried out within the framework of the project "Development of test-systems for early diagnosis of cardiovascular, oncological and neurodegenerative diseases based on miRNA associations and their target genes" № AP05132460 of the Ministry of education and science of the Republic of Kazakhstan (2018-2020).

Research approbation. The main results are presented and discussed:

- at the VII international scientific conference "Prospects for the development of biology, medicine and pharmacy" (Shymkent, Kazakhstan, 2018);
- at the international congress "Biotechology: state of the art and perspectives" (Moscow, Russia, 2019);

- at the fifth international conference "Plant genetics, genomics, bioinformatics and biotechnology" (Novosibirsk, Russia, 2019);

- at the VI international scientific conference of students and young scientists "Farabi alemi" (Almaty, Kazakhstan, 2019);

- at the 9th Moscow Conference on Computational Molecular Biology MCCMB'19 (Moscow, Russia, 2019).

- at the fifth international scientific conference "Current challenges in plant genetics, genomics, bioinformatics, and biotechnology" (Novosibirsk, Russia, 2019);

- at the international scientific conference of young scientists «Fundamental research and innovations in molecular biology, biotechnology, biochemistry» dedicated to the 80th anniversary of academician Murat Aitkhozhin (Almaty, Kazakhstan, 2019);

- at the international scientific and practical conference "The synergy of science and practice in the context of innovative breakthroughs in the development of economy and society: national and international aspects" (St. Petersburg, Russia, 2019);

- at the VII international scientific and practical conference "Biotechnology: science and practice" (Sevastopol, Russia, 2019);

- at the twelfth international multiconference "Bioinformatics of genome regulation and structure/systems biology" (Novosibirsk, Russia, 2020).

Publications.

The main content of the thesis is reflected in 16 printed works, including 2 articles in the international journal with the Impact Factor quoted in Scopus; 3 articles from the list of the Committee for Quality Assurance in Education and Science; 11 abstracts in materials of international conferences.

Author has conducted analysis of published data by investigated issue, the choice of objects of the study, the definition of aim and tasks, experimental studies and analysis of results of the research, writing and design of the thesis manuscript by herself.

Dissertation structure. This dissertation is written in 164 pages, and contains notations and abbreviations, introduction, literature review, materials and methods, results and discussions, conclusions, references and appendices from 422 sources, contains 32 tables, and 17 figures.