Theses of PhD dissertation

Dénes László Taller

KESZTHELY

2020

SZENT ISTVÁN UNIVERSITY

Comparative genome wide analysis of miRNAs in *Capsicum annuum*

DOI: 10.54598/000060

Dénes László Taller

KESZTHELY

2020

Comparative genome wide analysis of miRNAs in *Capsicum annuum*

The dissertation was prepared in order to earn a doctoral (PhD) degree at Georgikon Campus of the Szent István University, under the Plant Breeding, Genetics and Agricultural Biotechnology sub-program of the Festetics Doctoral School, in the discipline of Plant Breeding and Horticulture

By:
DÉNES LÁSZLÓ TALLER

Name of Doctoral School: Festetics Doctoral School
Head of school:

Dr. habil Angéla Anda, D.Sc.
professor
Szent István University-Georgikon
Department of Meteorology and Water Management

Supervisors:

Dr. habil János Taller, Ph.D

professor, head of department
Szent István University-Georgikon
Department of Plant Science and Biotechnology

Dr. Zoltán Havelda, D.Sc.
group leader, scientific advisor
NARIC-ABC
Plant Developmental Biology Group

1. Scientific background of the study

Nowadays, the recently discovered phenomenon of RNA interference and the associated non-coding RNA induced regulatory pathways are among the most popular biological research topics. In the classification of non-coding RNAs, microRNAs, due to their popularity, are often categorized in a separate class. MiRNAs are short, non-coding RNAs usually derived from endogenous genes (MIR genes) with their own promoters and are generally responsible for post-transcriptional inhibition of endogenous genes. The number of miRNA coding genes is in reach to the number of transcription factors. The miRNAs with high potential typically play a role in the regulation of essential functions or plant hormones and in the development of multiple tissues.

The required cost and time have been reduced significantly by the usage of next-generation deep sequencing technologies in the sequencing of genomes, non-coding RNAs and small RNAs, and the mass collection of RNA sequences became possible.

Many plant miRNAs have been identified experimentally or predicted by bioinformatical analysis of small RNA databases. A separate database was also established under the name MirBase to collect published miRNA sequences and annotations.

The class of conservative miRNAs identified in comparative studies of miRNA profiles of different plant species. Their common characteristic, that their sequence, target molecule, and function conserved during plant evolution. Many miRNAs can be characterized by an expression pattern specific to the plant family, species, tissue, or stage, or by expression change induced by environmental effects.

Comprehensive sequencing projects enrich our biological knowledge with a large amount of sequence data. However, they provide little information about the function of individual genes and small RNAs.

Other biotechnological methods are required to analyse and understand the functions associated with sequences.

Due to its speed, simplicity, and cost-effectiveness, the virus-induced gene silencing method (VIGS) has been preferred in many plant species, besides the cumbersome and costly solutions, that require stable genetic modifications. Creation of a specific VIGS vector usually done by the insertion of a part of the gene of interest into the genome of a well-characterized plant virus. The recombinant virus is able to infect the host plant, to spread in its cells, and to induce sequence-specific inhibition of target gene expression. Due to the induced RNA interference phenomenon, the target gene expression will drastically decrease and the specific knockdown phenotype will appear.

The VIGS experimental system always depends on the relationship between the chosen plant species and the used virus. When an existing vector starts to be used in a new species, tissue, or developmental stage, then the experimental system should be first tested with a reporter gene construct.

Peppers (*Capsicum annuum L.*) are one of the most important and widely grown vegetable crops. Previous studies of pepper fruit development have found significant tissue- and variety-specific differences in the metabolite levels and enzyme activity.

This process has been little studied at the level of mRNAs and even fewer at the level of miRNAs. The recently made available pepper genome greatly facilitates the processing and annotation of large amounts of RNA reads.

2. Objectives

1. The main objective of the research program is to set up a well-functioning assay system to identify new microRNAs and to provide detailed data on miRNA-level expression changes and interactions during pepper fruit development.

2. The methodological goal was to adapt the methods (RNA extraction, small RNA library construction, small RNA hybridization, virus-induced gene silencing) previously described and successfully applied on model plants (*Arabidopsis thaliana*, *Nicotiana benthamiana*) to pepper plants.

3. Using these adapted methods, we plan to determine the small RNA profiles of samples taken from paprika fruit tissues (flesh, seed, placenta) at different time points. The occurrence of conserved and pepper-specific miRNAs can be followed by our collected sequence data, and they can be characterized by bioinformatics methods. Based on the results of the sequencing, we can also find new, previously undescribed and most probably pepper fruit specific miRNAs.

4. Our aim is to involve individuals of both cultivated pepper (*C. annuum var. annuum*) and their wild ancestor (*C. annuum var. aviculare*), in our miRNA level studies and to determine their characteristic miRNA profiles. By comparing them in detail, we will be able to identify the new miRNAs that appeared in the cultivated version during domestication and the miRNAs that preserved exclusively in the wild species.

5. By comparative differential expression analysis of small RNA patterns of different tissues and developmental stages, our aim is to explore the significant miRNA differences, and the associated biological functions, and to identify tissue- or stage-specific expression of miRNAs.

3. Materials and methods

Plant material

We used the 'Fehér özön' pepper variety and 'Tepin' and 'Pequin' batches of Capsicum annuum var. aviculare for our miRNA-based assays. In the examined wild pepper batches we observed serious morphological differences, so we followed several traits to select a mother plant. Plants of 'Fehér özön' and 'Javított Bogyiszlói' cultivars were infected in our VIGS studies. TRV-VIGS constructs were passed onto pepper plants using the sap of infected *Nicotiana benthamiana* plants.

Elements of the virus induced gene silencing system

In vitro transcripts were generated and plants have been infected by the inoculation of 2-2 leaves for the testing of TMV and PVX-based vectors. For the TRV-VIGS vectors, the TRV1 and TRV2 constructs were multiplicated in the agrobacterial strain C58C1 and agroinfiltrated into the leaves of young Nicotiana benthamiana plants. After the onset of symptoms and the PDS reaction, the sap of selected tobacco leaves was spreaded on the carburundum-sprinkled leaves of our pepper plants.

RNA extraction and cDNA library preparation

RNAs were extracted from biological samples taken at 4 different time points after the time of anthesis using the TRI® Reagent RNA Isolation solution. This method reliably provided in sufficient concentration and pure RNA extracts from fruits, seeds and infected leaves. The quality of the extracted RNAs was paralelly checked by gel electrophoresis and spectrophotometer (NanoDrop 2000).

Selected high-quality RNA extracts were run on an acrylamide gel, and the small RNA region was marked on the gel and excised using a sc alpel. RNAs purified from the excised gel were used to prepare libraries. Small RNA libraries were generated using the Truseq Small RNA Sample Prep Kit (Illumina, CA, US) according to the manufacturer's instructions. The quality and concentration of the cDNA libraries were rechecked by gel

electrophoresis and the successful libraries were sent in bulks of 8-8 indexed samples for sequencing on the Illumina HiSeq 2000 platform.

Toolkit for bioinformatics analysis

Filtered pure miRNA reads were aligned to the genome *Capsicum annuum cv*. CM334. Conservative miRNAs were identified by using MirProf, this program searches the sequence-similarity between samples and the online database of miRNAs. For bioinformatics identification of new miRNAs, we used two (MirCat and miRDeep-P) software parallel. The merging of the two data sets and the necessary filtering steps were performed using Python scripts.

The Patman program was used to determine raw expression values. For differential expression studies, DESeq2 was used on the raw abundance matrix. MA diagrams and heat maps were generated to display the results of the differential expression analysis.

The PsRNATarget program was used to identify the target mRNAs of the examined miRNAs, and the annotated gene set of 'Zunla 1' variety was given as a reference gene list. The latest panther database was used for the ontological analysis of panther protein classes for genes and the GO Slim gene.

RNA-detection, RNA validation

Northern blot hybridization techniques were used to confirm the sequencing data, to determine the expression levels of endogenous plant genes, and to detect viral vectors.

Size separation of RNAs was required in all three cases, for each of them we used specific solutions. For small RNAs, we used 12% polyacrylamide gel containing 8M urea. A 1.2% agarose gel and around 100 V accelerating voltage were used to examine plants infected with VIGS vectors. In the study of endogenous genes, gel electrophoresis was performed on a 1.2% formaldehyde denaturing gel with 80 V accelerating voltage at 4 ° C.

RNAs were blotted from the gel to the membrane and after the required pretreatment steps hybridization was performed at the appropriate protocol temperature. After hybridization, unbound probes removed by several washing steps with different concentrations of SSC solution from the membrane. Then membranes were placed in an exposure cassette and the emitted radioactive signals were visualized using an X-ray film. The exposure time varied from a few minutes to several days depending on the intensity of the radioactive radiation measured on the membrane

Specific radiolabeled DNA or LNA oligos were used to detect miRNAs. A fragment of endogenous genes was amplified from the cDNA by PCR and ligated into a cloning vector. Plasmids isolated from selected bacterial colonies were also verified by restriction digestion and sequencing and subsequently used as templates for the multiplication of specific PCR products.

4. Results and discussion

Phenotyping

While the observation of the growing wild pepper accessions, we became aware of serious morphological differences, so we recorded 40 morphological traits and made a detailed description of each plant of these genetic materials. 11 'Tepin' plants were raised until ripening and a wild pepper specimen with the most ancient characteristics was selected for line formation.

Opportunities to use virus induced gene silencing in pepper

Our results showed that *Capsicum annuum* is particularly sensitive to infections with VIGS vectors. The PVX-based vector could not be used for function identification on peppers due to the caused severe necrotic symptoms. TMV and TRV infection brought milder symptoms and effective gene silencing, but both virus-based vectors induced drastic changes in the expression level of the tested reference genes. Therefore, in addition to qPCR-based assays, it is recommended to follow the induced changes of the gene silencing system by other hybridization-based methods.

Sequences obtained from small RNA libraries

Libraries were prepared from two biological replicates for each examined stage, tissue, and genotype. The sent 20 small RNA-libraries contained in total, more than 235 million reads.

The adapters, the invalid and low abundance sequences were excluded with several filtering steps from the 235 million raw sequences. We also excluded those sequences, which showed sequence-similarity with the entries of rRNA and tRNA database. 45.6% of the raw sequences met all the requirements provided by our screening steps. 95.43% of the obtained small RNA sequences were in the expected size range of 21 24 nt. The majority of sequences (52.5 percentage) were 24 nt in length, it is in line with those found in other species.

The size distribution of small RNAs showed significant differences between our samples, and we found for each tissue some specific characteristics in small RNA size distribution. In the flesh, the 21nt sequences showed a remarkably high rate. Based on these differences, it can be assumed that specific regulatory small RNAs of different sizes are involved in the developmental processes of each tissue.

Identification and classification of miRNAs

The filtered and genome-matched sequences from the 20 small RNA libraries were aligned to the miRBase database using the MirProf program. During this process, we detected match with 2033 miRNAs. After several filtering steps, we accepted a conservative miRNA list contained 217 unique sequences belonging to 40 conservative miRNA families.

We collected as capsicum-specific miRNA sequences, which were described as novel miRNAs in previous paprika studies. 68 unique sequences were listed, 42 were found in our 20 libraries by Patman alignment.

In 2017, another study was published (Liu et al., 2017) in which new miRNAs were described from pepper samples, ahead of our research group. This study contains 310 new miRNAs, mainly 24 nt in length, and shows only 1 sequence match with the sequences described in the previously mentioned studies. Compared to our data, we found four miRNAs included in both data sets. For the naming of these sequences, we took over their proposed IDs and moved these sequences from the list of new miRNAs to the group of pepperspecific miRNAs.

To discover previously undescribed, new miRNAs, we were able to identify 73 new potential miRNAs using a new identifying process, built from the parallel use of miRDeep-P and miRCat and the implementation of optimized filtering steps. For a better characterization, sequence lengths, starting nucleotides, and mean GC content were determined. Based on the sequence similarity within the new, predicted miRNAs, we were able to discover 6 new miRNA families with at least two members.

Validation of small RNA sequencing results

Validation tests of miRNA sequencing samples were examined in two groups: early-stage fruit samples and later-stage seeds, placenta and flesh samples. Small RNA northern blots and labelled DNA and LNA oligoprobes were used.

We found a large difference in the number of reads for certain miRNAs between the examined samples. By using the smallRNA Northern blot technique, quantitative differences could be detected for most of the tested miRNA expressions between the tested samples. Based on the sequenced reads, 36 of the examined miRNAs showed significant tissue-specific expression. For some of these miRNAs, the putative specific expression have also confirmed by hybridization. The results of the Northern hybridization studies were in good agreement with the sequencing data, although we also met with some significant differences.

Based on the comparison of the Northern hybridization and sequencing data, the amounts of small RNA reads per sample, collected by the high-throughput method, can be used for differential expression analysis of miRNAs during the development process. The deseq2-based differential expression assay was done in several approaches.

Differential expression analysis in the early stage of fruit development

For the early stages, we had the opportunity to compare the two genotypes and also the two stages.

For the comparison of two genotypes, we used parallel the absence-presence and deseq2 methods. The differential expression analysis was able to identify 25 unique miRNAs, which showed significant changes in expression levels at least in one stage between the two genotypes. In several miRNAs, the observed significant changes occurred at both stages, in addition, the direction of these changes was the same. Based on the two methods, we found new miRNAs in the cultivated variety

compared to the wild-type, while in the modern variety the expression and presumably functional role of certain conservative miRNAs were also reduced.

Comparing the two stages, we did not find any significant miRNA expression changes in Tepin. However, we found 10 variable miRNAs in the variety ‘Fehér özön’. In contrary to Tepin, it seems in 'Fehér özön', temporal changes in miRNA expression also play a role in miRNA-based regulation.

Differential expression analysis in the late stage of fruit development

MiRNAs often show a tissue-specific expression pattern, so samples from different tissues of the crop were compared to explore these changes at both the 28- and 40- DAA stages. In our investigations, we found an exceptionally high number of miRNAs that showed elevated expression levels in the seeds but we were able to identify some miRNAs, which can be characterized by flesh and placenta-specific expression changes.

Between the two chosen time points, all three examined fruit tissues underwent significant morphological changes. These changes are accompanied by the alteration in the expression of the miRNAs. We observed significant changes in expression of 17 miRNAs in the flesh and 19 miRNAs in the seeds. However, in the placenta, no miRNA changes were found in the comparison of the two mentioned time points, which can be counted as significant..

The mRNAs, genes and functions regulated by identified miRNAs

We successfully used PsRNATarget to identify a large number of potential target mRNAs for the examined conservative, pepper-specific, and novel miRNAs. Among the predicted target-miRNA relationships has been found several that previous studies in tomatoes or peppers have identified as validated or putative miRNA-target relationships.

5. Conclusions and Future Prospects

In this study, comprehensive, genome-level, and tissue-specific information was collected for the small RNA expression pattern of the pepper fruit development. In the flesh, the observed remarkably high expression rate of 21 nucleotide long small RNA molecules support, the fact, that they play an important role in pepper fruit development.

The detailed description of conservative miRNA profiles provides the opportunity to compare the identified miRNA sequences and their expression levels and regulatory functions with related species.

A remarkable number of pepper-specific and novel miRNAs were identified in our samples. The description of the new miRNAs, their expression profile and their precursors can not only help to understand the molecular background of crop development and to refine existing developmental biological and domestication models, but the obtained information can be used directly to develop new, miRNA-based, biotechnological methods, specifically for the target crop-, stage- or tissue.

We were able to assign potential precursors and star strand sequences to each of the identified miRNAs. This linked sequence information eases to molecularly investigate the maturation and biogenesis of miRNAs from selected precursors.

By annotating the miRNA regulated mRNAs, we can get an overview, those biological processes and regulatory cycles, where the miRNAs play a decisive role. The complex information gathered during the work allows the selection of miRNAs and target mRNAs involved in specific processes. In addition, by incorporating the selected genes into the TRV vector construct presented in this dissertation, it is possible to perform functional studies of the chosen mRNAs even inside the pepper fruit.

6. New scientific results

I. **I created 20 small RNA sequencing libraries from pepper fruits and different fruit tissues. From these libraries, 217 conservative miRNAs, and 42 previously described pepper miRNAs were identified, and several were confirmed with hybridization techniques.**

II. **I observed that different tissues (placenta, seed, flesh) have different small RNA size distribution. The 21 nt long sRNAs occurred at the highest rate in the flesh samples, especially at 40 DAA.**

III. **A new miRNA prediction pipeline was developed by combining two existing bioinformatic tool (miRDeep P and miRCat) with some tested, additional filtering steps. With this system, I was able to identify 73 new, reliable, individual miRNA sequences with their precursors and star-strands from pepper fruits.**

IV. **With testing the absence and presence in the two tested genotypes ('Tepin'and 'Fehér özön') 6 miRNA was identified 'Fehér özön' specific, I found it in every stage and tissue of it, while have not found in any sample from 'Tepin'. From this six, five has been validated by differential expression analysis (deseq2).**

V. **In the differential expression analysis, I identified miRNAs with seed-specific (17), flesh specific (11) and placenta specific (2) expression changes, which was significant in both stages, and it was also significant compared to both other tissue, from the same stage. We found 3 miRNA, which expression changed drastically between the two time-points.**

VI. **With the testing of 3 different viruses (TMV, PVX, TRV) based VIGS constructs I realized, that Capsicum annuum is really sensitive to VIGS infections. VIGS can have different effects on pepper plants, depends on the used virus, but in all tested cases visible symptoms emerged and the expression of usually used qPCR reference genes altered drastically.**

VII. **I used TRV VIGS vector constructs for N. benthamiana agroinfiltration, and with the mechanical transfer, I was able to infect pepper plants with high efficiency. In some cases, these constructs were able to provide functional gene silencing also in the fruits of the infected pepper plants.**

7. List of publications

Presentations in English

Dénes Taller, Jeannette Bálint, Tibor Nagy, Endre Barta, Éva Várallyay, Zoltán Havelda: Genome wide investigation of differential expression of sweet pepper (Capsicum annuum) micro RNAs during fruit formation Hungarian Molecular Life Sciences 2015., 27-29. March 2015. , Eger

Bálint, Jeannette ; Gyula, Péter ; Taller, Dénes ; Dalmadi, Ágnes ; Hamar, Éva ; Kis, András ; Szittya, György ; Várallyay, Éva ; Taller, János ; Havelda, Zoltán (2017): Investigation of the regulation and activity of RNA interference executor complexes in model and crop plants In: Hungarian Molecular Life Sciences 2017 (2017) p. 171

Articles in english in professional journals with impact factor

Taller D, Bálint J, Gyula P, Nagy T, Barta E, Baksa I, et al. (2018) Expansion of Capsicum annuum fruit is linked to dynamic tissue-specific differential expression of miRNA and siRNA profiles. PLoS ONE 13(7): e0200207.

Taller, D., Bálint, J., Gyula, P., Nagy, T., Barta, E., Baksa, I., ... & Havelda, Z. (2018). Correction: Expansion of *Capsicum annum* fruit is linked to dynamic tissue-specific differential expression of miRNA and siRNA profiles. PloS one, 13(8), e0203582.

Enikő Oláh, Réka Pesti, Dénes Taller, Zoltán Havelda, Éva Várallyay (2016): Non-targeted effects of virus-induced gene silencing vectors on host endogenous gene expression. Archives of Virology (2016) 161:2387–2393 DOI 10.1007/s00705-016-2921-9

Hegedűs A.*, Taller D.*, Papp N., Szikriszt B., Ercisli S., Halász J., Stefanovits-Bányai É. (2013): Fruit antioxidant capacity and self-incompatibility genotype of Ukrainian sweet cherry (Prunus avium L.) cultivars highlight their breeding prospects Euphytica, Vol. 191, Issue 1, pp 153-164.

Presentations in Hungarian

Taller D., Bálint J., Nagy T., Várallyay É., Havelda Z., MikroRNS-ek azonosítása a paprika (*Capsicum annuum*) termésfejlődése során, PSAK XII. Pro Scientia Aranyérmesek XII. konferenciája. Absztraktkötet. pp: 35.

Taller D., Bálint J., Nagy T., Várallyay É., Havelda Z., *Capsicum annuum* szabályozó kisRNS-ek azonosítása új generációs szekvenálással, MBK napok 2014, Gödöllő

Taller D., Bálint J., Nagy T., Taller J., Barta E., Várallyay É., Havelda Z., Terméspecifikus mikroRNS-ek azonosítása paprikában, Magyar Növénybiológiai Társaság Fiatal Növénybiológusok 2015 Konferenciája, 2015. február 6. Pécs

Taller D., Bálint J. Havelda Z., A paprika genomszintű megismerése új generációs szekvenálási módszerekkel, Innovatív Kertészeti Kutatások a Pannon Régióban című MTA PAB Kertészeti Munkabizottsága tudományos ülése, 2014. május 14., Keszthely

Taller D. Az RNS-ek titokzatos világa, célkeresztben a paprika mikroRNS-ek! Festetics Imre Mezőgazdasági Biotechnológia Szakkollégium, 2013. 12.03., Gödöllő.

Oláh E, Pesti R, Taller D, Várallyay É, Havelda Z, Fiatal Biotechnológusok Országos Konferenciája 2014 március 7., Szeged

Oláh E., Pesti R., Taller D., Havelda Z., Várallyay É. (2014): Vannak-e a VIGS-nek nem kívánt mellékhatásai? (2014) előadás: MBK Napok, 2014, Gödöllő, 2014. 12. 11.

Full length abstracts of Hungarian presentations

Taller Dénes, Bálint Jeannette, Nagy Tibor, Várallyay Éva, Havelda Zoltán, MikroRNS-ek azonosítása a paprika *(Capsicum annuum)* termésfejlődése során, PSAK XII. Pro Scientia Aranyérmesek XII. konferenciája. Tanulmánykötet. Líceum Kiadó 2015. pp: 134-138.

Articles in Hungarian in professional journals

Bálint Jeannette , Kis András , Taller Dénes , Nagy Tibor , Barta Endre , Molnár János, Tusnády E. Gábor, Marincs Ferenc és Havelda Zoltán.(2015), Az RNS-interferencia szerepe a növények patogénekkel szembeni védekezésében és a fejlődésbiológiai folyamatokban, Növényvédelem 51 (12), 2015 p. 539-549.

Articles in Hungarian in other journals

Oláh E, Taller D, Várallyay É, Vírusok a biológia szolgálatában, TermészetBúvár 2014/5 12-14.