Final Report K 109438

Genome wide mRNA and small RNA transcriptome profiling and characterization in Capsicum annuum.

NKFI-OTKA K 109438 2013-09-01 - 2017-11-30

Principal investigator: Zoltan Havelda

Introduction

The discovery that eukaryotic organisms, including plants, employ several RNA interference (RNAi) pathways in which small regulatory (s)RNAs (20 to 24 nucleotide in length) mediate the negative regulation of their target RNAs in a sequence-specific manner broadened our understanding of gene regulation. In plants, RNAi pathways play indispensable roles in developmental processes and in biotic and abiotic stress responses. Distinct RNAi pathways generate different classes of small regulatory RNAs such as microRNAs (miRNAs) and small interfering RNAs (siRNAs).

The continuously increasing number of experimental studies are demonstrating that plant tissue and organ differentiation is regulated by specific sRNAs. Scientific experiments proved the importance of miRNAs and tasiRNAs in many aspects of plant development, including leaf, shoot, flower and root development, phase-change, auxin signalling and response to environmental/biotic stresses. In addition to the well-known evolutionarily conserved miRNA families, it is extremely important to identify and characterise new, non-conserved miRNAs that are present in closely related species, lineage-specific miRNAs. The identification of new miRNAs and the genome-wide, tissue-specific analyses of miRNA expression can help to specify miRNA-regulated developmental processes substantially affecting the quality of economically important traits of crop plants.

The role of miRNAs in fruit development has been extensively investigated by NGS technology. Several studies analysed the expression of miRNAs at a genome-wide level during fruit development in tomato, a Solanaceae model plant. Fruit development in another important Solanaceae species, the pepper (Capsicum annuum), is less extensively investigated. Conserved and potentially new, species-specific miRNAs have been described as a result of an investigation of different pepper organs including the developing fruit as a whole. However, since the activity of miRNAs is spatiotemporally highly regulated it is inevitable to get detailed tissue-specific information about the expression pattern of miRNAs during fruit development. Moreover, little is known about the contribution of sRNA-mediated regulation to the fleshy fruit development at the phase of cell expansion where the economically valuable fruit volume is often to reach a two magnitude fold increase.

MiRNAs have been revealed as a crucial component in fine regulation of gene expression during developmental processes. MiRNAs modulate gene expression by regulating the degradation or translational repression of their target mRNAs. Fully functional mature miRNAs are generated from their precursors by some phylogenetically conserved protein families in the following manner: the nuclear genome-coded miRNA genes (MIR) are transcribed by RNA polymerase II (Pol II) to produce primary miRNAs (pri-miRNAs) possessing a specific and stable hairpin-like secondary stem-loop structure. In the nucleus, an RNAse III enzyme, DICER-LIKE1 (DCL1) cleaves the hairpin out from the rest of the pri-miRNA resulting in the production of 80–300-nt-long miRNA precursors (pre-miRNAs). In plants, the same DCL1 enzyme then cleaves the precursor at sites determined by structural features to generate small double-stranded miRNA intermediates, the miRNA:miRNA* duplexes. These duplexes are transported to the cytoplasm where they are loaded into the RNA-induced silencing complex (RISC). This complex displaces one of the strands (miRNA*) while the mature strand (miRNA) remains bound. The miRNA-containing RISC mediates sequence-specific repression via mRNA-cleavage or translational repression. The main executor of RISCs are the ARGONAUTE (AGO) proteins. The Arabidopsis thaliana genome encodes ten AGO proteins which are specialised for various RNAi pathways but often display functional redundancies. AGO1 is the central component of the RISC playing dominant role in miRNA-mediated regulation and its expression is regulated by a feedback mechanism-mediated by miR168.

Beside miRNAs there is a less defined group of hairpin-associated RNAs that cannot be classified as miRNAs (other hairpin RNA, ohpRNA;). This class of sRNAs are numerous, and are mainly derived from inverted repeat elements that are transcribed into long RNA (retrotransposons). They are characterised by longer precursors, larger, more complex loops and less precise maturation than miRNAs. However, they resemble miRNAs in their stranded production (the self-complementary single-stranded precursor folds back to form double-stranded stretches that are recognised by various Dicer-like endonucleases) and that usually there are one or two major RNA species which are much more abundant than the other sequences produced from the same precursor. From ohpRNA loci predominantly 24-nt sRNAs are produced. The function of this class of sRNA is not known, it is presumed that similarly to heterochromatic siRNAs (hetsiRNAs) they are involved in the repression of transposable elements. The generation of siRNAs is different from that of miRNAs because it is dependent on an RNA-dependent RNA polymerase (RDRP) [3]. The perfect double-stranded (ds)RNA intermediates generated by specific RDRPs can be recognised and cleaved by Dicer-like enzymes other than DCL1 to produce various classes of siRNAs. HetsiRNAs for example are mainly 24-nt in length and generated by DCL3 via processing of Pol IV/RDR2-dependent dsRNA species originating from intergenic or repetitive regions of the genome. These 24-nt-long siRNAs are loaded into the RNA-induced transcriptional silencing (RITS) complex, which is functionally different from the RISC. They recognise Pol V-dependent short transcripts emerging from repetitive and intergenic regions and recruit chromatin-modifying complexes. HetsiRNAs play fundamental roles in maintaining genome integrity, by initiating de novo DNA methylation on transposable and repetitive elements mainly during reproductive transitions (meiosis, gametogenesis and embryogenesis), especially in plants with larger, repetitive genomes. Phased (pha)siRNAs, however, are derived from Pol II-dependent mRNAs which are cleaved by a special class of miRNAs/siRNAs of 22-nt length. This cleavage renders one of the cleavage products to be a substrate of RDR6/SUPPRESSOR OF GENE SILENCING 3 (SGS3). The resulting dsRNA is then cleaved by DCL4 endonuclease following a distinctive, regular (phased) pattern starting from the cleavage site. The phasiRNAs are preferentially loaded into a distinctive RISC containing AGO5.

Results

NGS (next generation sequencing) experiments

Pepper small RNA and mRNA sequencing experiments from total RNA samples deriving from developing fruit.

ID	Sample	Geotype	days (DAA)	tissue
A	**D110_10+11**	Fehér özön 5	40	pulp
B	**D111_8-10m**	Fehér özön 5	40	pulp
C	**D111_5m**	Fehér özön 5	40	seed
D	**D110_18-19m**	Fehér özön 5	40	seed
E	**D110_20+21**	Fehér özön 5	40	placenta
F	**D111_6-7**	Fehér özön 5	40	placenta
G	**D115_11-16**	Fehér özön 5	28	pulp
H	**D111_21-24**	Fehér özön 5	28	pulp
I	**D107_9+10**	Fehér özön 5	28	seed
J	**D113_11-12**	Fehér özön 5	28	seed
K	**D111_18-20**	Fehér özön 5	28	placenta
L	**D115_8-10**	Fehér özön 5	28	placenta
M	**D108_2-3m**	Tepin6	19	fruit mix
N	**D108_4**	Tepin6	20	fruit mix
O	**D108*_3-4-6m**	Fehér özön 5	19	fruit mix
P	**D108_11-12m**	Fehér özön 5	20	fruit mix

Table 1. List of pepper samples used for small RNA and mRNA NGS

Investigation of pepper RNA samples at 28 and 40 DAA

To get a high-resolution spatiotemporal information about dynamic sRNA expression pattern associated with pepper fruit development in the expansion phase fruit was dissected to get flesh, seed and placenta samples at two time points, 28 and 40 days after anthesis (DAA), and investigated by NGS methodology. The mRNA and sRNA sequencing were performed on an Illumina HiSeq2000 platform by UDGenomed, Debrecen. Using the gained datasets, we generated the comparative tissue-specific differential expression profiles of the conserved, already described and our newly predicted pepper-specific miRNAs. We found that differential expression activity of miRNAs are intriguingly high at 40 DAA and especially the expansion of flesh is associated with profound changes in the miRNA expression profile. The pronounced role of miRNA-mediated regulation in flesh development was also reflected by the high level of AGO1 protein present in flesh samples. Moreover, the analyses of non-miRNA-like sRNAs revealed an ample of pepper-specific hetsiRNAs and phasiRNAs playing a role mainly in the development of the seed and the placenta. As a summary, our work demonstrates that the cell expansion phase of the economically important sweet pepper is under intensive sRNA-mediated regulation implying that breeding technologies capable of modifying the expression of key sRNA regulators can be potent tools to enhance the economic value of pepper.

Based on these data a manuscript prepared for submission titled,

“Expansion of Capsicum annum fruit is linked to dynamic tissue-specific differential expression of miRNA and siRNA profiles Dénes Taller,Jeannette Bálint1,Péter Gyula, Tibor Nagy, Endre Barta, Ivett Baksa, György Szittya, János Taller and **Zoltán Havelda”** see as attached document.

We also investigated the mRNAseq data form this experiment and the preliminary data presented on Figure 1. Gene expressions were measured in all the three tissues at two time points and compared to each other. Just as in the case of pepper-specific and novel miRNAs (and sRNA loci) expressions, the least changes were observable in the placenta and the flesh between 28 and 40 DAA. Also, there were only a few changing expressions the placenta and the flesh at 28 DAA.

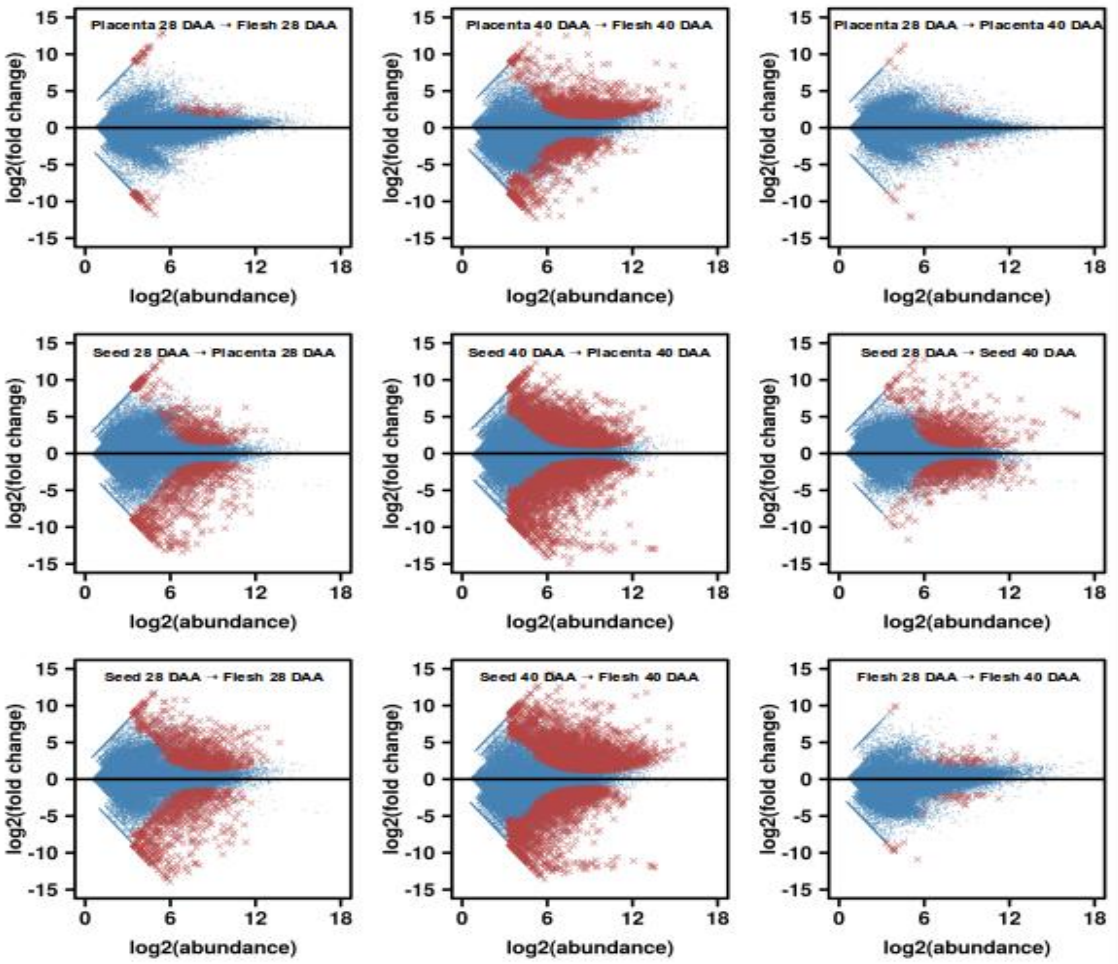


Figure 1. Differentially expressed mRNA transcripts in the pepper tissues at two developmental stages.

Bioinformatic analysis of the sRNA and mRNA expression patterns in pepper cultivars, Tepin, Fehérözön and dihaploid (DH) Fehérözön.

In the next experiments we compared the sRNA and mRNA profiles of the modern cultivar Fehérözön and the botanical wild type ancestor of the cultivated pepper species, Tepin, which has small fruit at 14 and 20 DAA. These experiments may allow to reveal the changes in sRNA and mRNA profiles during the formation of the economically important fruit of the pepper. We also incorporated samples from dihaploid (DH) Fehérözön, DH lines exhibit extreme homozygosity and are extensively used in breeding programs. We were interested whether this property of the genome is associated with changes in RNA profiles. The 50 bp single-end reads were processed to remove adaptors and to filter low quality sequences. The sRNA

sequences were further filtered to keep only the 20-25 nt range. Next, the rRNA- and tRNA-derived sequences were removed. The cleaned sRNA sequences were aligned to the *Capsicum annuum* v1.55 genome using ShortStack 3.8.4 (Johnson *et al.*, 2016). First, sRNA loci were predicted with ShortStack, using an abundance cutoff of 1 RPM (read-per-million). In this way, 145689 loci were identified and their raw abundances (read counts) were determined. Normalization and differential expression analysis was performed using DESeq2 (Love *et al.*, 2014). We considered a locus differentially expressed if the fold-change was at least 2, the normalized abundance was at least 100, and the adjusted p-value (by the Benjamini-Hochberg method) was less than 0.01 (1% false positive hit). The expression of the sRNA loci were graphically represented on MA-plots, where the expression fold-changes are plotted against their average abundance (Figure 2 and 3).

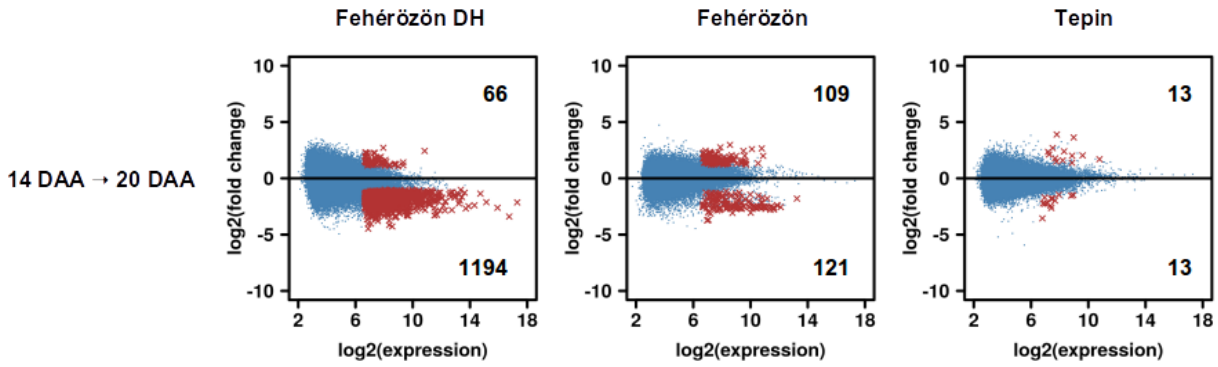


Figure 2. Differentially expressed sRNA loci between 14 and 20 DAA (days after anthesis). The highest number of changed loci between 14 and 20 DAA was observed in the Fehérözön dihaploid samples (1194 down-regulated and 66 up-regulated loci). In the wild pepper (Tepin), only a few loci changed. There are no loci that were changed in all the 3 pepper variant but there were many common between Fehérözön and the Fehérözön dihaploid, as expected.

Many of the down-regulated loci in the dihaploid pepper are expressing 21/22-nt siRNAs in a phased manner, suggesting that some protein coding or non-coding genes are down-regulated in the dihaploid pepper at the later stage of development but this regulation is missing from the Fehérözön or the Tepin. The majority of the differentially expressed loci in

the dihaploid pepper are 24-nt siRNA producing loci.

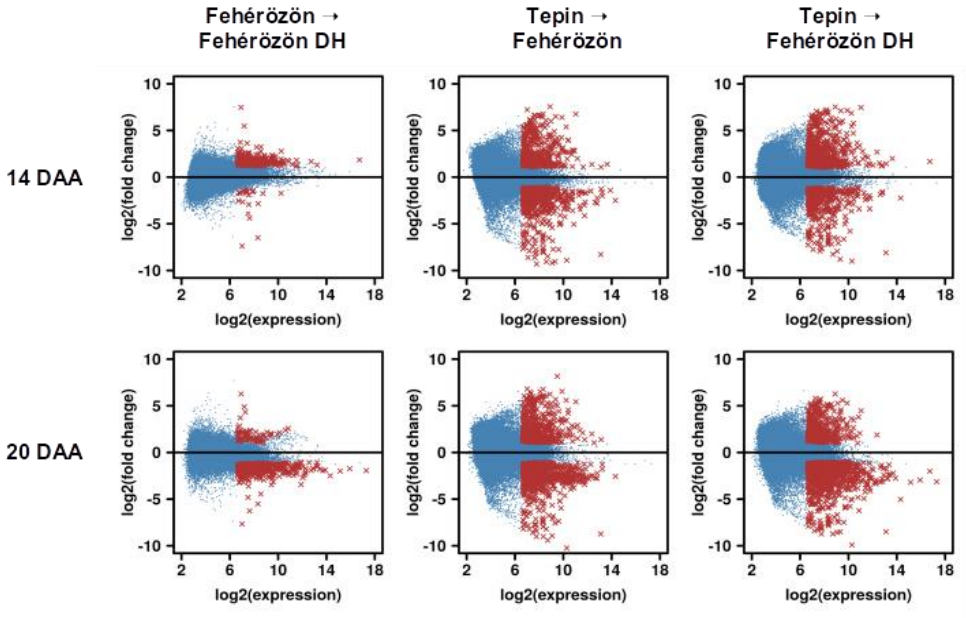


Figure 3. Differentially expressed sRNA loci between pepper variants at 14 and 20 DAA (days after anthesis). The highest number of changed loci can be observed between the Tepin and the Fehérözön or dihalpoid Fehérözön at both time points. Interestingly, there are many differentially expressed loci between Fehérözön and dihaploid Fehérözön, suggesting that in these apparently closely related variants the fruit development differs, at least at the small RNA level.

After loci prediction, sequences were studied as functional units. Genome-matching non-redundant sequences were collected and counted in all the samples. Counts were normalized to library sizes and the normalized abundances were visualized on heat maps (Figure. 4). Sequences were annotated by matching them with the known and predicted miRNA sequences, miRNA precursors, mRNAs from the Zunla pepper, RNAs assembled by us, and the transposon sequences of the Zunla pepper. According to this, there were Tepin- or Fehérözön-specific sequences that mostly matched LTR retrotransposons, receptor kinase genes (resistance genes) or not matching any annotated genes, suggesting that they are intergenic in origin. When we analysed the miRNA sequences, it turned out that most of them are present in all pepper variants but we found three miRNAs (can-mir-n030, can-mir-n018 and can-mir-f13) that were apparently missing from Tepin samples (Figure 5).

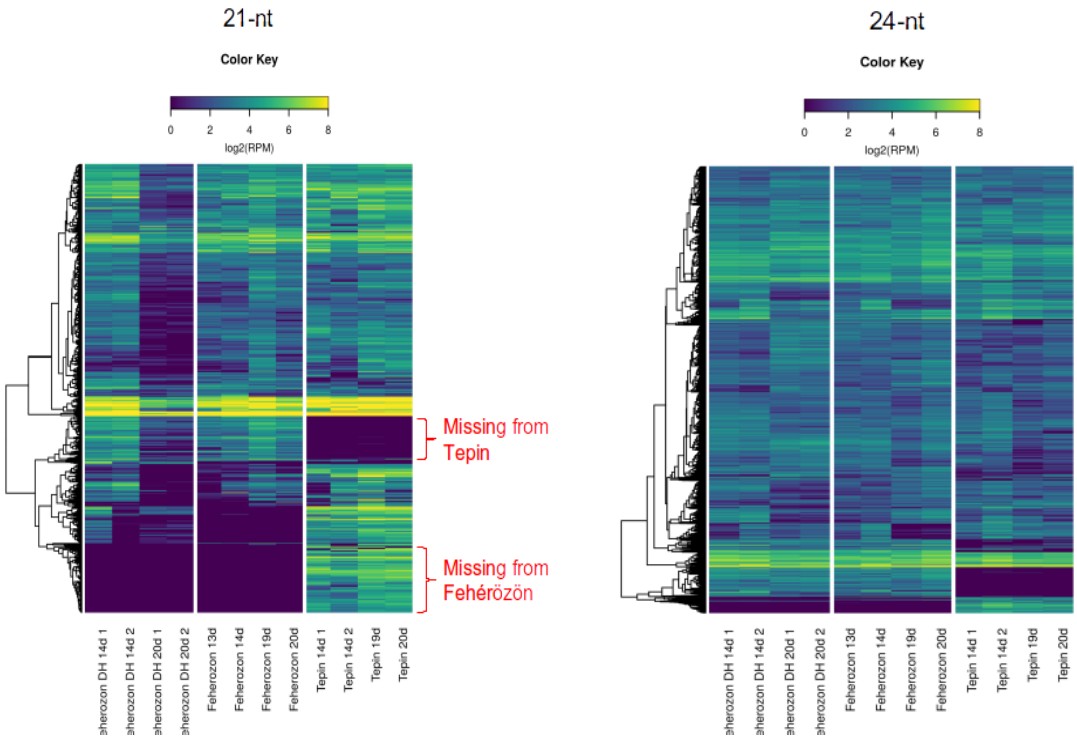
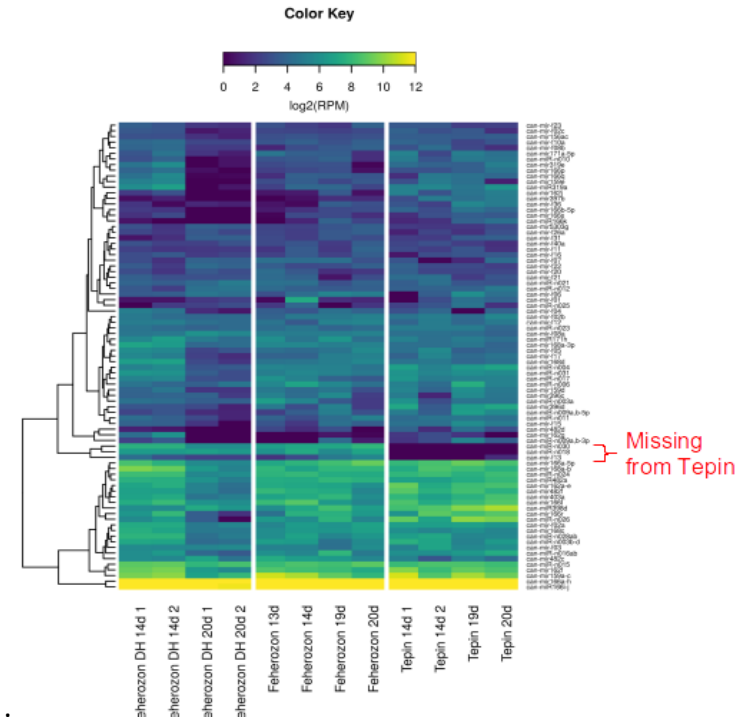


Figure 4. Heat maps showing the expression of 21-nt and 24-nt sRNA loci. There are Tepin- and Fehérözön-specific sequences.



.

Figure 5. Heat map showing the expression of identified miRNA sequences.

Most of the miRNA sequences are present in all the samples, except three miRNAs, that are missing from Tepin. The can-miR-n030 (22-nt) and can-miR-n018 (24-nt) were identified by others as pepper-specific miRNAs, while can-mir-f13 was identified by us as pepper-specific. This suggests that these miRNAs may be linked to some developmental processes that are Fehérözön-specific. The predicted targets of these miRNAs are listed in Table 2.

miRNA	Expectation	Target gene	Target annotation
can-miR-n030	1.5	XLOC_042819	Capsicum annuum putative disease resistance protein At1g50180 (LOC107859122)
can-miR-n030	1.5	XLOC_015993	Capsicum annuum putative disease resistance protein At1g50180 (LOC107859122)
can-miR-n030	3.5	XLOC_001864	Capsicum annuum alcohol dehydrogenase-like 7 (LOC107849669)
can-mir-f13	0	XLOC_016418	Capsicum annuum uncharacterized LOC107841160 (LOC107841160)
can-mir-f13	2.5	CA04g22070	Capsicum annuum la-related protein 1C-like (LOC107866855)
can-mir-f13	3	CA07g07670	Capsicum annuum serine/threonine-protein kinase HT1-like (LOC107877872)
can-mir-f13	3.5	CA12g06700	Capsicum annuum uncharacterized LOC107850394 (LOC107850394)
can-mir-f13	4.5	XLOC_045422	Capsicum annuum PLASMODESMATA CALLOSE-BINDING PROTEIN 3-like (LOC107847866)

Table 2. Predicted targets of the Fehérözön-specific miRNAs.

Targets were predicted with psRNATarget server. Only hits with predicted cleavage were retained. There was no predicted target for can-mir-n018, which is a 24-nt miRNA. This miRNA might actually be a hairpin-associated siRNA that usually targets DNA-methylation instead of mRNA cleavage.

For the RNA-seq analysis (mRNA profile), clean reads were aligned to the previously assembled pepper transcriptome sequences (including known and novel alternative transcript variants) using kallisto, the differential expression analysis was performed with sleuth (Pimentel *et al.*, 2017). We filtered for those transcripts whose normalised expression was at least 10, their fold change were more than 2, and the adjusted p-value was lower than 0.01 (1% false positive hit). We found that plenty of transcripts changed during development (between 14 and 20 DAA) in all the pepper variants (Figure 6).

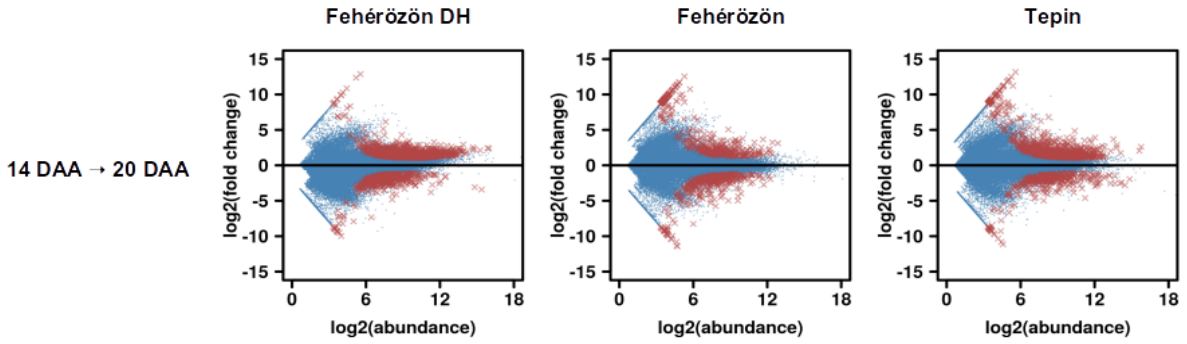


Figure 6. Differentially expressed transcripts between 14 and 20 DAA (days after anthesis). Many transcripts changed during fruit development in all the 3 pepper variants but only 13 (0.4%) changed in all the 3. Only 71 (3%) changed both in Fehérözön and the dihaploid Fehérözön.

Surprisingly, only a small fraction (3%) of the differentially expressed transcripts were common between the Fehérözön and the dihaploid Fehérözön. One of the most down-regulated transcript that changes only in the dihaploid Fehérözön, is a GAGA-binding transcription factor (Figure 7) that regulates gene expression at multiple level. It is highly sensitive to chromatin-modifications, therefore this difference may indicate a markedly changed chromatin state in the dihaploid pepper compared to the parental Fehérözön.

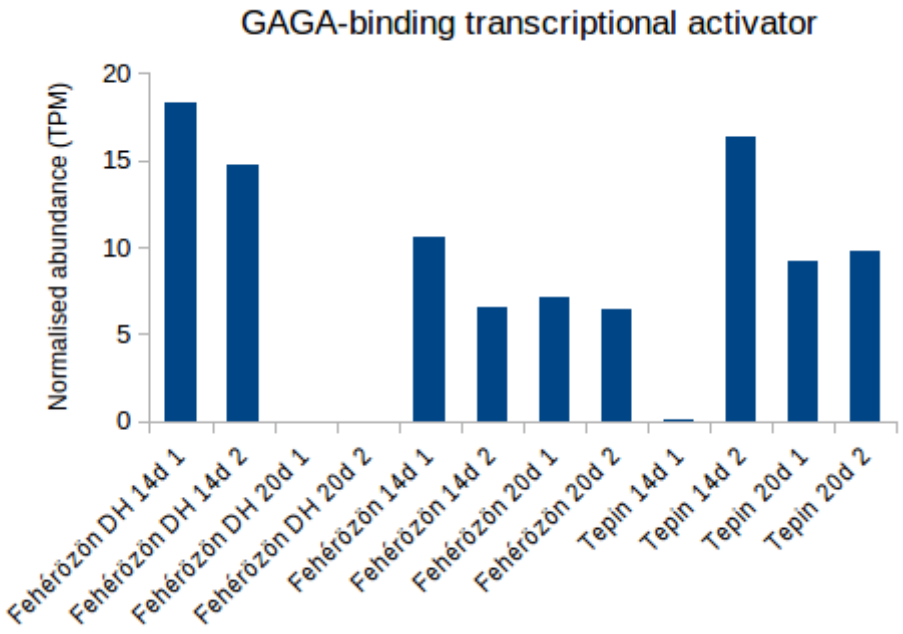


Figure 7. Expression of a GAGA-binding transcription factor (CA04g23120).

GAGA-binding transcription factors are very sensitive to the chromatin state, therefore the apparent lack of expression may indicate an altered chromatin state.

There are apparently Tepin-specific genes, that can be important from an agricultural point of view. For example, when we compare Tepin with Fehérözön or dihaploid Fehérözön, one of the most prominent difference is the expression of a Kunitz-type protease inhibitor (Figure 8).

These protease inhibitors are important for resistance but may have an adverse effect on the human digestion apparat.

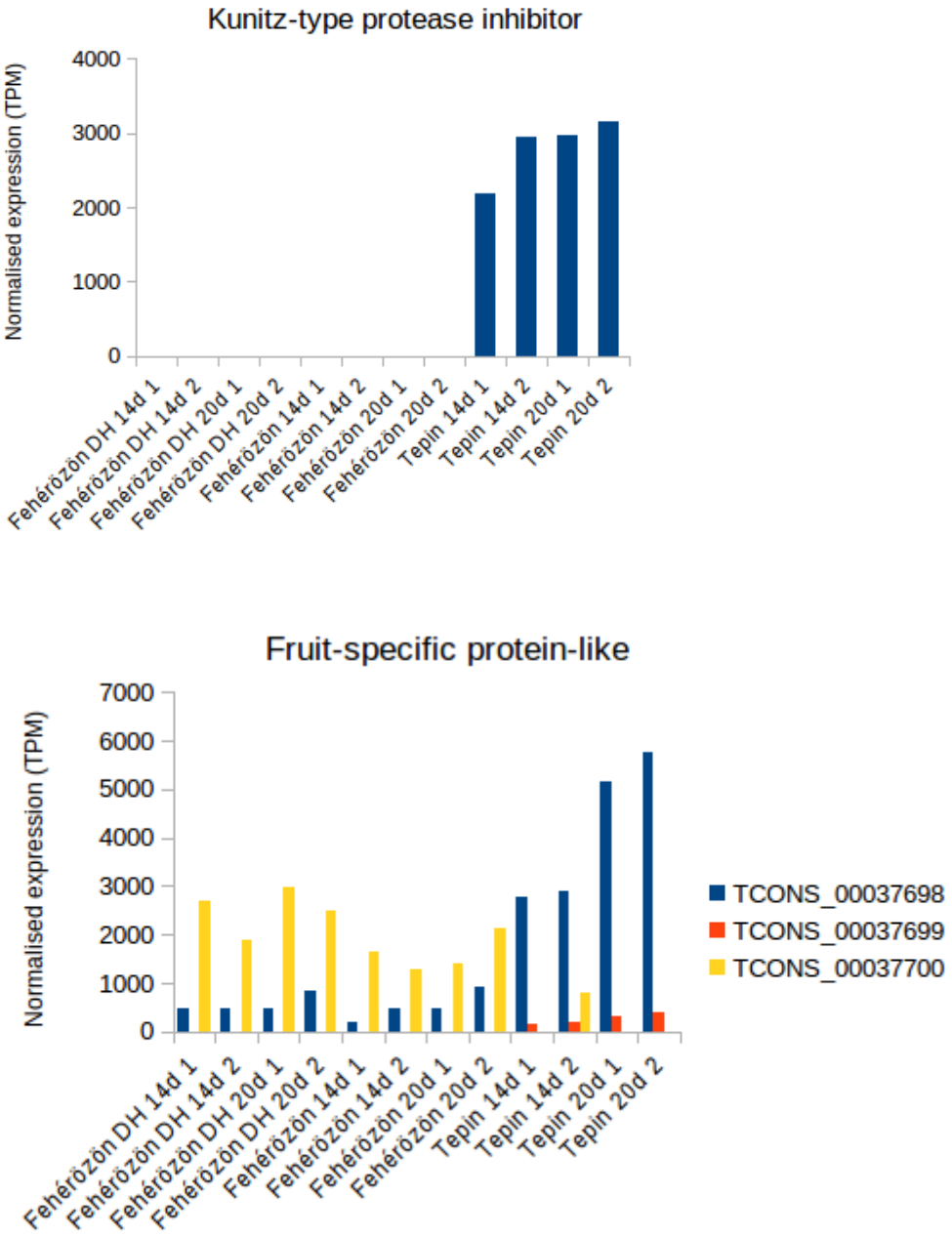


Figure 8. Expression of highly differentially expressed genes between Tepin and Fehérözön.

A Kunitz-type protease inhibitor (CA02g09720) was found to be exclusively expressed in Tepin (Figure 8). These kind of protease inhibitors are important for the plant to keep away herbivores. However, it might not be a desired trait in the case of a cultivated, edible plant. Another gene (XLOC_018289), however, apparently has alternative transcripts that are highly

expressed in the cultivated pepper, while another is more expressed in the wild type plant Figure 8). The function of this fruit-specific gene is unknown. The lack of expression of this gene might indicate a possible positive selection for a deletion event in the Fehérözön cultivar by humans. There are other missing protease inhibitors (e.g. cystein proteases), lectins and other genes that can be associated to the cultivation process. On the other hand, there are genes that are more expressed in the cultivated pepper. A fruit-specific protein coding gene with three alternative transcripts shows an interesting expression pattern. One transcript is more expressed in the Tepin, whereas an alternative transcript is almost exclusively expressed in the Fehérözön and dihaploid Fehérözön samples. The function of this gene is unknown. Altogether, there are 3823 differentially expressed genes between Tepin and Fehérözön.

Next, we compared the expression of genes at one time point between the pepper cultivars. In accordance with that there is a small number (3%) of common DEGs between the Fehérözön and the dihaploid Fehérözön between 14 and 20 DAA, plenty of genes were found to be differentially expressed between the two variants but only at the later time point (Figure 9). This, together with the finding that the sRNA profiles are also markedly different between these two closely related pepper variants suggests that their fruit developmental program runs differently. Considering the method by which the dihaploid pepper is obtained, it is highly likely that the altered chromatin state (i.e. DNA methylation) might be responsible for the differences. A thorough investigation of the differentially expressed genes may clear the picture.

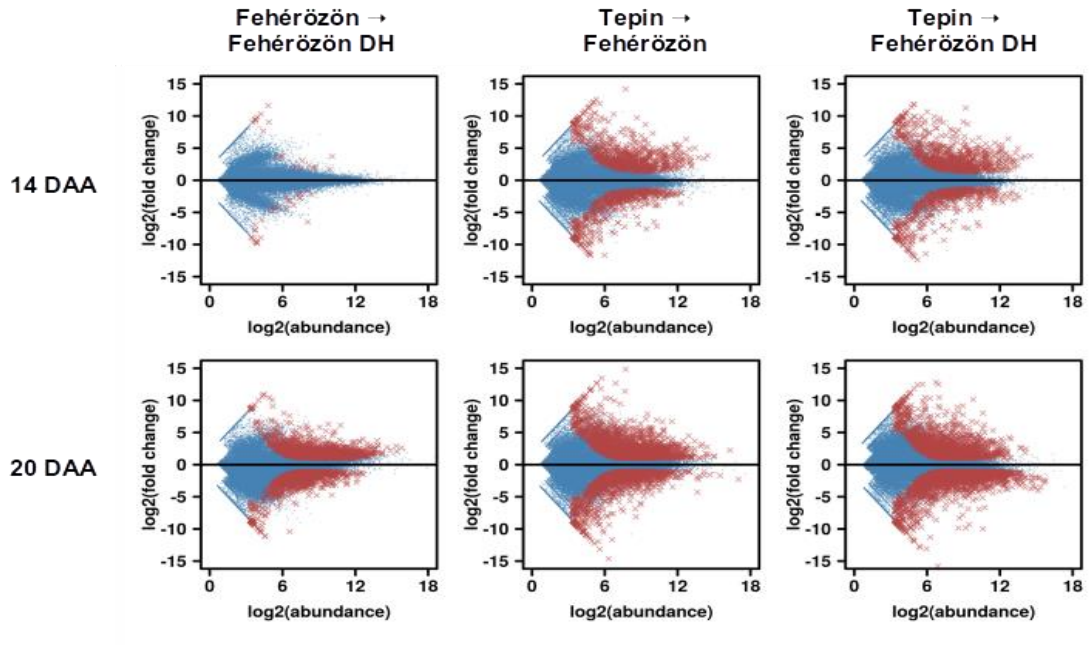


Figure 9. Differentially expressed transcripts between pepper cultivars. Gene expression were compared between pepper cultivars to reveal major differences in their developmental programs. Surprisingly, there is a huge difference between the Fehérözön and the dihaploid Fehérözön at the later developmental stage but not at the early stage.

Investigation of sRNA landscape by gel-filtration experiments.

Our NGS experiments revealed massive amount of small RNA species present in the plants. However, our wet lab experiments also revealed that there are at least on miRNA (miR168) which is present in protein unbound form. These data indicated that small RNAs sequenced form total RNA extracts not necessarily reflects the biological activity of the sequenced small RNAs. To facilitate to identification of biological relevant sRNAs, incorporated into protein complexes, we used a method based on gel filtration (Superdex 200) to separate protein complexes which contain different small RNA species. These experiments will allow the identification of small RNAs incorporated in executor complexes. First we used Arabidopsis thaliana model plant to test this experimental approach. Young leaf and flower crude extracts were applied on gel filtration column and collected fraction were used small RNA northern blot experiments (Figure 10)

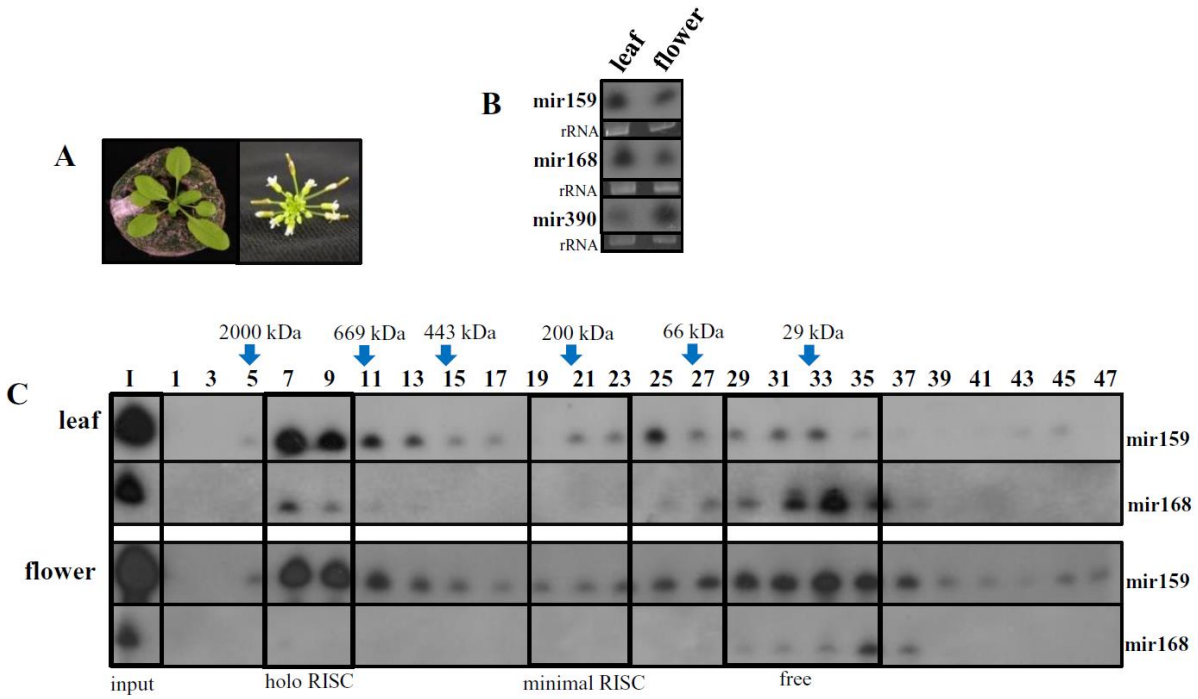


Figure 10. Experimental scheme and molecule size separation with FPLC. A *Arabidopsis thaliana* Columbia wild type leaf and flower. B Small RNA hybridization from leaf and flower total RNA extract. We detected conservative miRNAs (mir159; mir168 and mir390) with LNA probes from both of type of samples. C: Small RNA hybridization from the gel filtration experiments. 1 to 47 show the each RNA fractions. We detected the same conservative miRNAs and defined three fraction parts based on the protein markers. Input: total RNAs, holo RISC: Argonaute and other protein bound sRNAs; minimal RISC: Argonaute bound sRNAs; free: unbound sRNAs. Blue arrows show the size of protein markers. Black boxes show those fractions from which we made small RNA libraries.

Based on these experiments we identified three categories small RNAs, bound to high molecular weight complexes (holo RISC, biologically active), small RNA present at about 200 KDa (possibly incorporated into single AGO protein, minimal RISC, biologically active) and a biologically inactive category at about 29 KDa which represents protein unbound sRNAs. Next, using purified RNAs from the fractions we prepared small RNA libraries which were used for NGS (Table 3). We also made western blots to detect the main executor proteins and successfully identified AGO4 in the minimal RISC fractions (Figure 11).

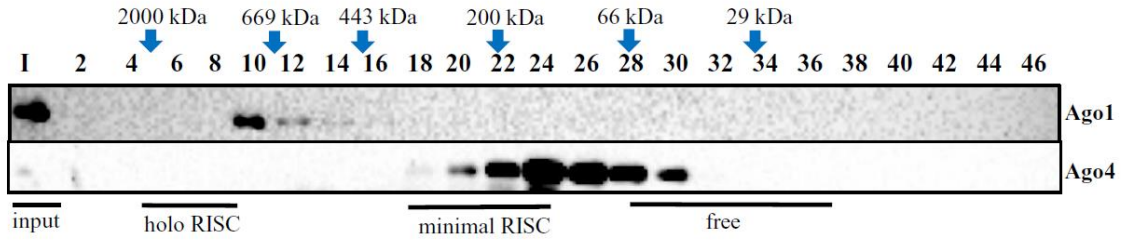


Figure 11. Western hybridization for Argonaute proteins Blue arrows show the size of protein markers.

Small RNA library ID	Sample	Index	Tissue	Species
J187li	J187 leaf input	A/1	Leaf	*Arabidopsis thaliana* wt.
J188li	J188 leaf input biological repeat	A/2	Leaf	*Arabidopsis thaliana* wt.
J187lk	J187 leaf holo RISC	A/4	Leaf	*Arabidopsis thaliana* wt.
J188lk	J188 leaf holo RISC biological repeat	A/5	Leaf	*Arabidopsis thaliana* wt.
J187lm	J187 leaf minimal RISC	A/6	Leaf	*Arabidopsis thaliana* wt.
J188lm	J188 leaf minimal RISC biological repeat	A/7	Leaf	*Arabidopsis thaliana* wt.
J187lsz	J187 leaf free	C/25	Leaf	*Arabidopsis thaliana* wt.
J198lsz	J198 leaf free biological repeat	C/26	Leaf	*Arabidopsis thaliana* wt.
J190vi	J190 flower input	A/3	Flower	*Arabidopsis thaliana* wt.
J191vi	J191 flower input biological repeat	A/4	Flower	*Arabidopsis thaliana* wt.
J190vk	J190 flower holo RISC	A/5	Flower	*Arabidopsis thaliana* wt.
J191vk	J191 flower holo RISC biological repeat	A/6	Flower	*Arabidopsis thaliana* wt.
J190vm	J190 flower minimal RISC	A/7	Flower	*Arabidopsis thaliana* wt.
J191vm	J191 flower minimal RISC biological repeat	A/8	Flower	*Arabidopsis thaliana* wt.
J190vsz	J190 flower free	C/27	Flower	*Arabidopsis thaliana* wt.
J191vsz	J191 flower free biological repeat	C/28	Flower	*Arabidopsis thaliana* wt.

Table 3. Samples used for small RNA libraray preparation for NGS.

The NGS data of the fractions revealed that there are many sRNAs, which are not incorporated into the AGO-RISC and present in free, protein unbound form in the cell (Figure 12).

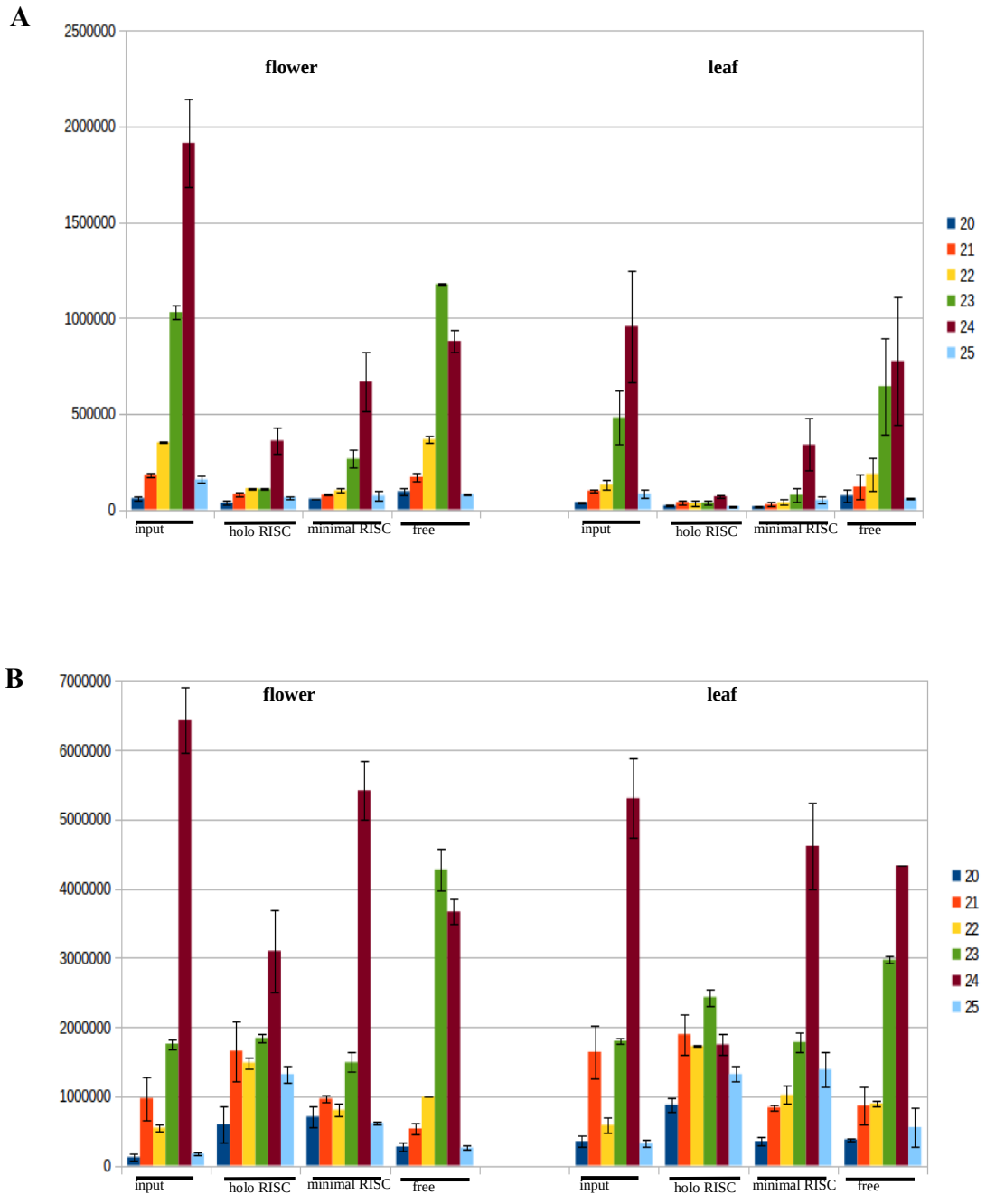


Figure 12. Size distribution of non-redundant (A) and redundant (B) normalised sequences in the defined fractions.

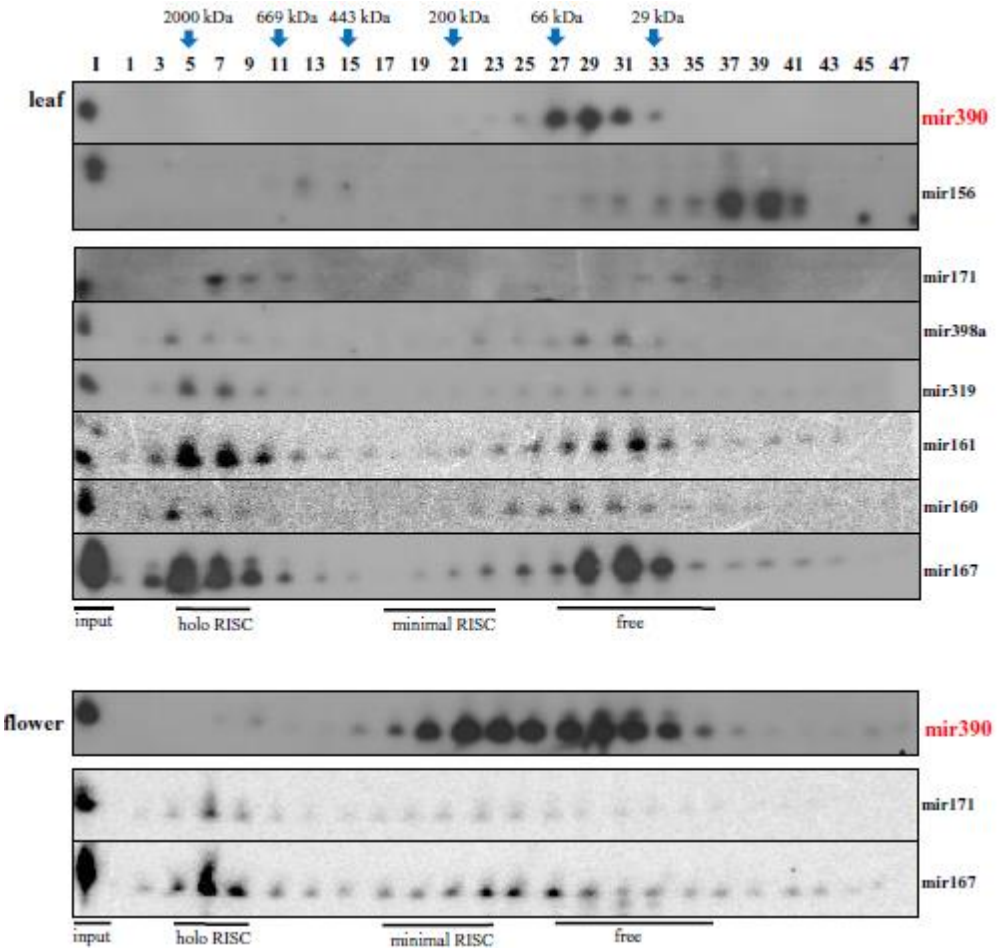


Figure 13. Conservative miRNA Northern hybridization for RNA fractions

These data indicate that the investigation of RISC loaded sRNAs will facilitate the identification of biologically relevant sRNAs in the investigated samples. Next generation sequencing of total RNA, RISC and protein bound sRNAs revealed that sRNA content of the cell is sorted in different manners. Our results suggest that a portion of sRNAs sequenced from total RNA sample is not active and the biological relevance of a given sRNA in the sample is determined by its presence in the executor proteins. The bioinformatics data were validated by northern hybridization experiments where fractions of the gel infiltration were applied on membranes (Figure 13). The hybridization experiments confirmed the bioinformatics data. So we were able to detect holo RISC and minimal RISC bound sRNAs and ample of free, protein unbound sRNA species. Since, our data on the model plant, were proved to be reliable we used this

methodical approach in pepper. To gain high resolution information on pepper fruit development we used dissected pepper fruit at 28 DAA to apply crude extracts of seed, placenta and flesh on Sephacryl S300 gel filtration column. We optimized the protocol for pepper and carried out the experiments in two independent biological repeats. The NGS sequenced samples are listed on Table 4.

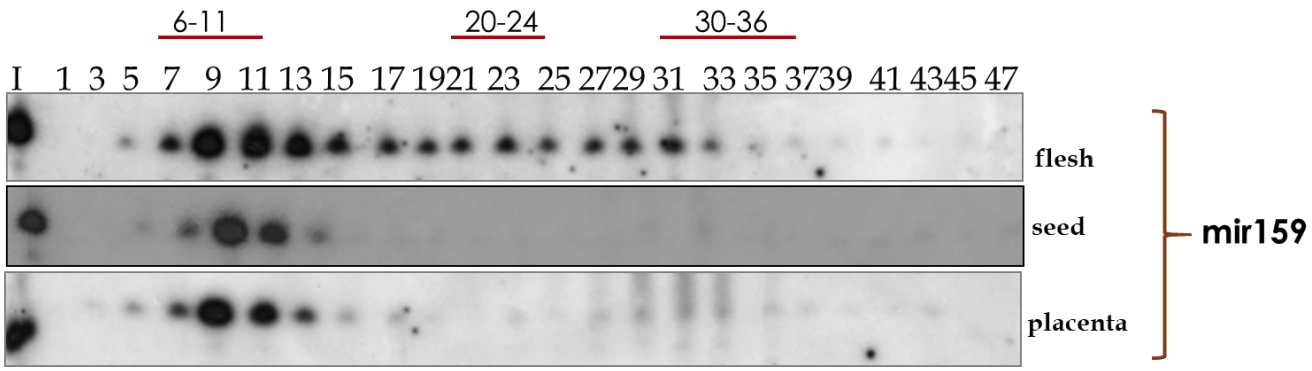


Figure 13. miR159 Northern hybridization for RNA fractions prepared from pepper fruits flesh, seed and placenta.

The preliminary investigation of size distribution revealed, that similarly to Arabidopsis gel-filtration data, RISC bound fractions show different size distribution in all tissue types compared to the input sample (Figure 14). This indicates that sequencing of total RNA extract not necessarily reflects the biologically active pool of small RNAs. Moreover, our data also suggest that detailed analyses of RISC bound small RNAs can help to identify new still unknown sRNA species or biologically important sRNAs characteristic for the particular developmental stage. Heat map analyses of the expression pattern of abundant small RNAs (Figure 15) also underlies the importance of gel filtration based small RNA analyses since drastic differences can be detected in between the different samples deriving RISC bound, free or total RNA extracts.

Small RNA library ID	Sample	Index	Tissue	Species
J211 SEED i	J211 Fö528 seed input	A/1	Seed	*Capsicum annuum* „Fehérözön”
J224 SEED i rep.	J224 Fö528 seed input biological repeat	A/2	Seed	*Capsicum annuum* „Fehérözön”
J211 SEED k	J211 Fö528 seed holo RISC 6-11	A/1	Seed	*Capsicum annuum* „Fehérözön”
J224 SEED k rep.	J224 Fö528 seed holo RISC biological repeat	A/2	Seed	*Capsicum annuum* „Fehérözön”
J211 SEED m	J211 Fö528 seed minimal RISC 20-24	C/29	Seed	*Capsicum annuum* „Fehérözön”
J224 SEED m rep.	J224 Fö528 seed minimal RISC biological repeat	C/30	Seed	*Capsicum annuum* „Fehérözön”
J211 SEED f	J211 Fö528 seed free 30-36	C/25	Seed	*Capsicum annuum* „Fehérözön”
J224 SEED f rep.	J224 Fö528 seed free biological repeat	C/26	Seed	*Capsicum annuum* „Fehérözön”
J214 PLACENTA i	J214 Fö528 placenta input	A/3	Placenta	*Capsicum annuum* „Fehérözön”
J225 PLACENTA i rep.	J225 Fö528 placenta input biological repeat	A/4	Placenta	*Capsicum annuum* „Fehérözön”
J214 PLACENTA k	J214 Fö528 placenta holo RISC 6-11	A/5	Placenta	*Capsicum annuum* „Fehérözön”
J225 PLACENTA k rep.	J225 Fö528 placenta holo RISC biological repeat	A/6	Placenta	*Capsicum annuum* „Fehérözön”
J214 PLACENTA m	J214 Fö528 placenta minimal RISC 20-24	A/7	Placenta	*Capsicum annuum* „Fehérözön”
J225 PLACENTA m rep.	J225 Fö528 placenta minimal RISC biological repeat	A/8	Placenta	*Capsicum annuum* „Fehérözön”
J214 PLACENTA f	J214 Fö528 placenta free 30-36	C/27	Placenta	*Capsicum annuum* „Fehérözön”
J225 PLACENTA f rep.	J225 Fö528 placenta free biological repeat	C/28	Placenta	*Capsicum annuum* „Fehérözön”
J215 FLESH i	J215 Fö528 flesh input	A/3	Flesh	*Capsicum annuum* „Fehérözön”
J226 FLESH i rep.	J226 Fö528 flesh input biological repeat	A/4	Flesh	*Capsicum annuum* „Fehérözön”
J215 FLESH k	J215 Fö5 flesh holo RISC 6-11	A/5	Flesh	*Capsicum annuum* „Fehérözön”
J226 FLESH k rep.	J226 Fö528 flesh holo RISC biological repeat	A/6	Flesh	*Capsicum annuum* „Fehérözön”
J215 FLESH m	J215 Fö528 flesh minimal RISC 20-24	A/7	Flesh	*Capsicum annuum* „Fehérözön”
J226 FLESH m rep.	J226 Fö528 flesh minimal RISC biological repeat	A/8	Flesh	*Capsicum annuum* „Fehérözön”
J215 FLESH f	J215 Fö528 flesh free 30-36	C/27	Flesh	*Capsicum annuum* „Fehérözön”
J226 FLESH f rep.	J226 Fö5 flesh free biological repeat	C/28	Flesh	*Capsicum annuum* „Fehérözön”

Table 4. Pepper samples used for small RNA libraray preparation for NGS.

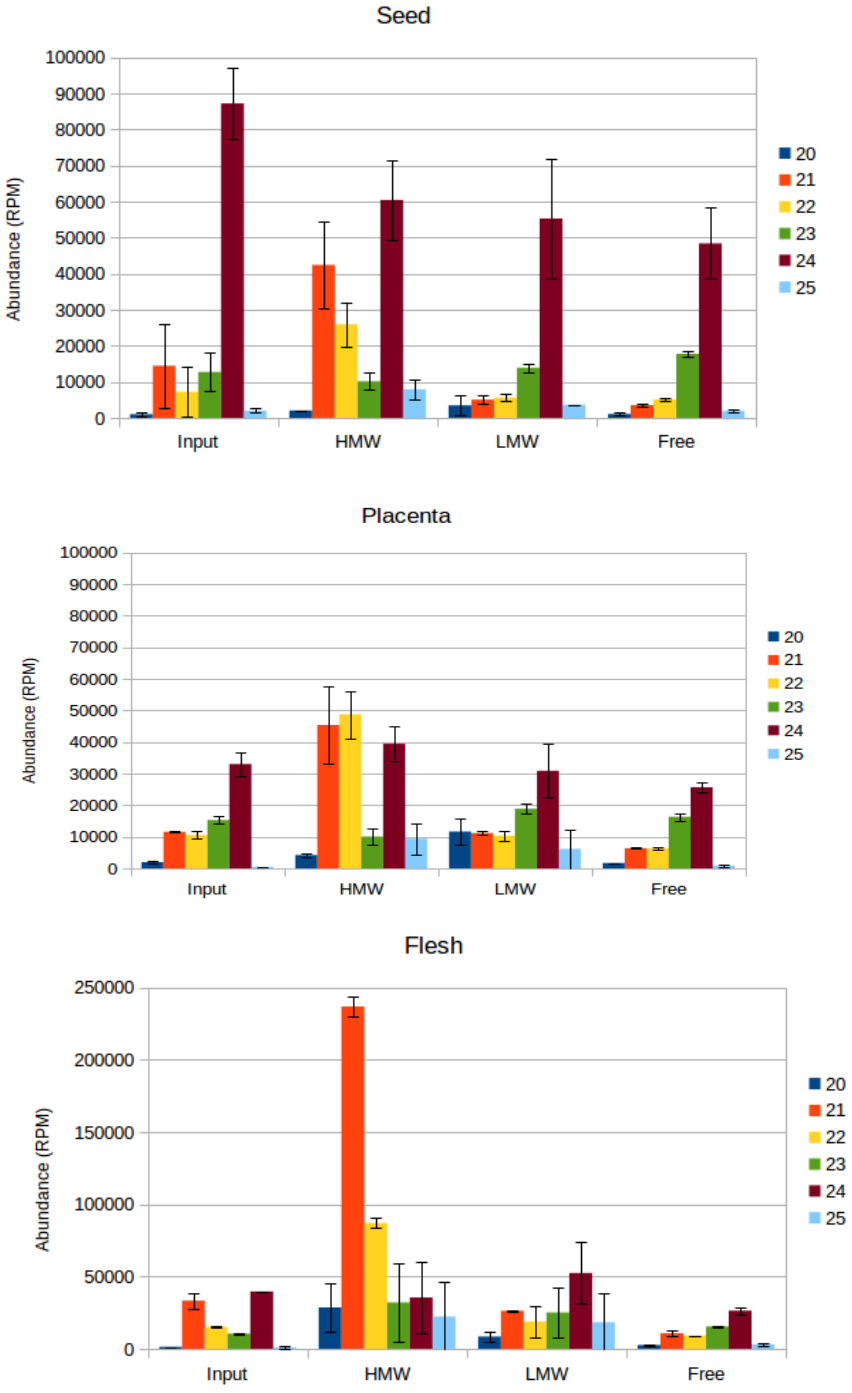


Figure 14. Size distribution of normalised sequences, abundancy higher than 5, in the defined fractions and input samples in different pepper tissues.

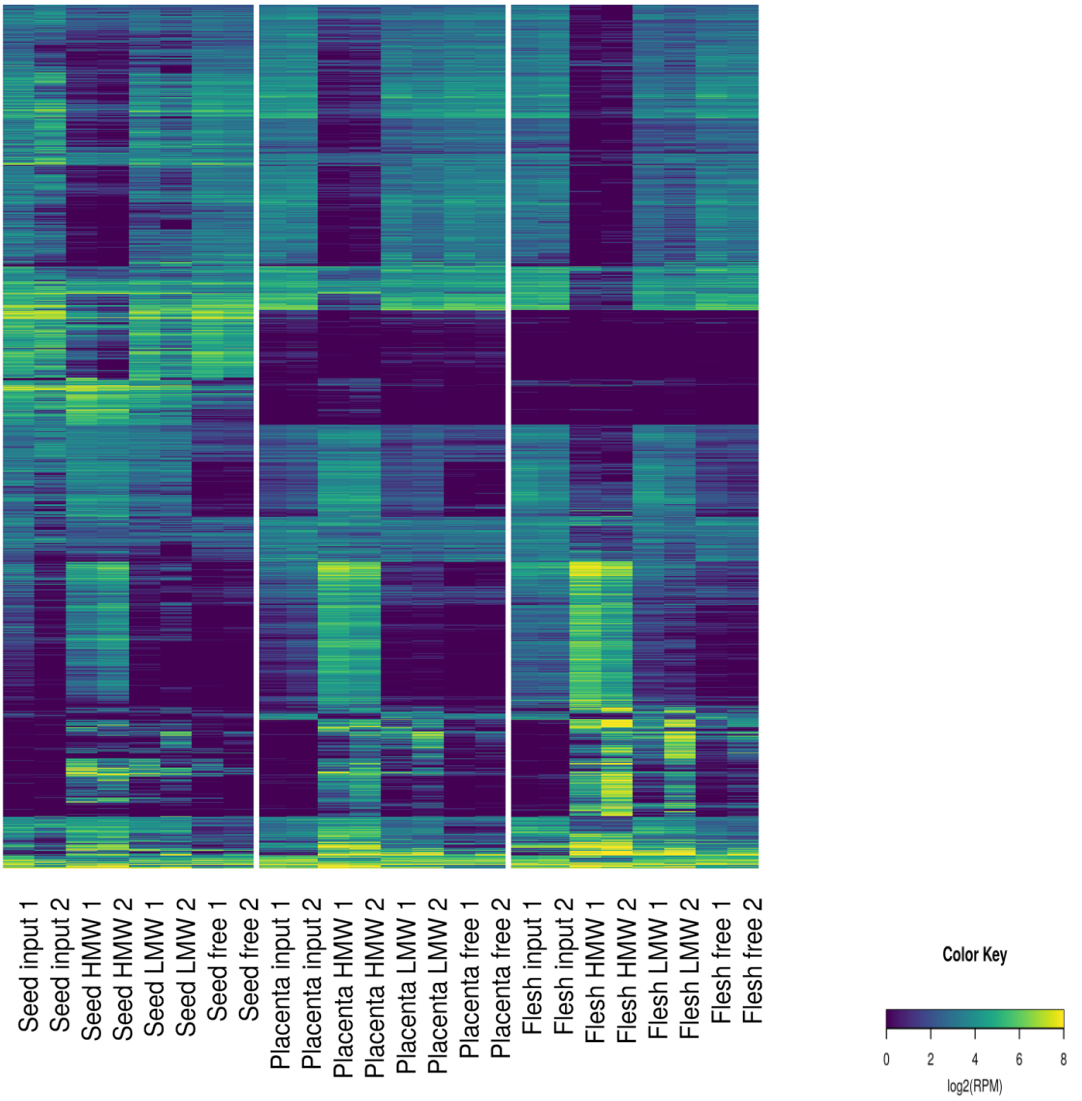


Figure 15. Heat map showing the expression of identified sRNA sequences in different pepper fractions and in input total RNA extract.

Technological improvements

Virus induced gene silencing (VIGS) on Capsicum annuum

Since genetic transformation of pepper is technically demanding or not possible depending on the genotype, a transient system for the modulation of target RNA levels would be very useful. We had preliminary data about the usage of VIGS in pepper and we further investigated the potential of this technology. PVX infection causes very severe symptoms and necrosis on pepper, consequently this virus cannot be used as a VIGS vector for this plant. C. annuum

plants were infected with TMV-VIGS and TRV-VIGS vectors and their effect was analysed. Both empty and PDS-containing VIGS vectors induced similar changes. Northern blot analyses revealed that TMV infection induced drastic down-regulation of Rubisco and Gapdh mRNAs. Unfortunately our attempts to make a Northern blot with Ef and Actin mRNAs failed. Interestingly, contrary to our results on other hosts, in TRV-VIGS infected pepper the expression levels of Rubisco and Gapdh mRNAs were up-regulated. To explain this phenomenon we have to make further investigations. As a summary we could not find an optimal vector either for pepper and our data show that using VIGS on pepper requires extra attention since the used virus vectors can induce drastic alterations in the host gene expression.

These results have been published in "Oláh E, Pesti R, Taller D, **Havelda Z,** Várallyay É. (2016) Non-targeted effects of virus-induced gene silencing vectors on host endogenous gene expression. Arch Virol. 2016 Sep;161(9):2387-93. IF 2.2"

Genetic transformation of pepper

To test the biological function of the identified target RNAs an efficient protocol for pepper transformation would be necessary. However, pepper is highly recalcitrant for agrobacterium-mediated transformation and the transformation protocols are inefficient and highly genotype dependent. The main bottleneck is the inability to generate pepper cells that are competent for Agrobacterium infection and have the ability to regenerate. We have attempted to establish a transformation protocol for the cv. Fehérözön used in our experiments (Figure 16). To achieve this goal we have tried to transform a expression cassette expressing BABY BOOM (BBM) AP2/ERF transcription factor, since the expression of this protein has been described to facilitate the efficient regeneration of transgenic sweet pepper. The establishment of cv. Fehérözön line expressing BBM would be the base for over-transformation experiments, where over-expression or knocking-out of the candidate mRNA or miRNA targets would allow the investigation of their biological functions. We have started the transformation experiments using different agrobacterium species and tissue culture conditions. To date we were able to achieve usable regeneration rate of pepper callus but our attempts to produce transgenic Fehérözön failed. Further experiments are in progress and the transformation parameters are in the focus of recent attempts.

Figure 16. Tissue culture steps for plant regeneration from Fehérözön callus and callus generation from Agrobacterium infected leaf bits.

Artificial (a)miRNA technology

An alternative approach to investigate role of target RNAs (RNAs or miRNA precursors) the application of artificial miRNA (amiRNA) mediating the degradation of the selected target RNAs. AmiRNAs are designed by modifying the miRNA/miRNA* duplexes of an endogenous miRNA precursor to produce an artificial miRNA targeting experimentally defined target sequences. The advantage of this technology is that the precise miRNA processing facilitates to assess and hence reduce the potential off-target effects of the expressed small RNA in the donor plant. Moreover, since the miRNA pathway is an indispensable regulatory system governing developmental processes it is expected that low temperature does not affect its activity. We identified a very compact conservative miRNA (miR171) in barley plant. The precursor of this miRNA is relatively short and contained a restriction site which facilitated its modification with biotechnological approaches. With inserted an extra restriction site in the precursor, retaining the structural features, for easy manipulation. As a result this precursor cab be turn into artificial miRNA very easily using only to oligonucleotides. We demonstrated that, in addition to its extreme power to restrict the accumulation of a virus in barley plants, it works very efficiently in transient assays in Nicotiana benthamina indicating that this amiR backbone will be suitable also in Solanaceae plants.

These results was published in "Kis A, Tholt G, Ivanics M, Várallyay É, Jenes B, **Havelda Z**. (2016) Polycistronic artificial miRNA-mediated resistance to Wheat dwarf virus in barley is highly efficient at low temperature. **Mol Plant Pathol**. 17. (3):427-37. IF 4.7"

Nicotiana benthamiana miRNA NGS

N. benthamiana is a Solenaceae plant widely used model plant species for research on plant-pathogen interactions as well as other areas of plant science. To have the possibility to use N benthamiana small RNA NGS data as cross reference in our experiments we took part in the NGS experiments of this species. In this study Illumina sequencing was used to identify small RNAs, their cleaved target mRNAs followed by Northern blot validation of the miRNAs. 40 conserved and 18 novel microRNAs and validated their target mRNAs were identified in *N. benthamiana* with a genomic scale method. The targets were found to be involved in various biological processes including transcription, RNA binding, DNA modification, signal transduction, stress response and metabolic processes. In further experiments we will bweable to use this data in comparison with pepper data. These data was published in, Baksa I, Nagy T, Barta E, **Havelda Z**, Várallyay É, Silhavy D, Burgyán J, Szittya G. (2015) Identification of Nicotiana benthamiana microRNAs and their targets using high throughput sequencing and degradome analysis. **BMC Genomics**. 16:1025. (IF 3.8)

NGS library preparation

NGS techniques are increasingly used to identify and discover small RNAs. For these sequencing methods, preparation of libraries from isolated nucleic acids is required. The library preparation can be performed using either total RNA or the isolated sRNA fraction. We compared data obtained by sequencing small RNA libraries prepared from both the isolated total RNA and the sRNA fraction. We demonstrated that sequencing the libraries prepared from isolated small RNA fractions and from total RNA resulted in very similar numbers of raw reads, but the absolute and normalized values of both the raw and processed data had higher variability among the libraries made from isolated small RNA than from total RNA. This finding indicate that isolation of small RNA fraction before library preparation is very advantageous.

This observation was published in " Nagy T, Kis A, Poliska S, Barta E, **Havelda Z,** Marincs F. (2016) Comparison of small RNA next-generation sequencing with and without isolation of small RNA fraction. Biotechniques. Jun 1;60(6):273-8. (IF 2.9)"

Summary

The aim of our work to analyse small (s)RNA and mRNA profiles of pepper during fruit formation to understand this fundamental, and economically also important, developmental process. The original work plan aimed the partial sequence analyses of pepper genome, however several pepper genomes have been published superseding our plan. Using high-throughput sequencing (next generation sequencing; NGS) we determined the small and mRNA profiles of a modern cultivar and wild type pepper at several fruit developmental stages. To gain high resolution tissue specific information we dissected the fruit for placenta, seed and flesh at later developmental stages. We demonstrated with comparative differential expression analyses of conserved, already described and our newly predicted pepper-specific miRNAs revealed that fruit expansion is accompanied by an increasing level of sRNA-mediated regulation of gene expression. We also elaborated a size separation chromatography based identification, instead of degradom analyses, of biologically active sRNA species. This methodological approach has been applied and optimized for different pepper tissues (flesh, seed and placenta). Small RNA content of the identified protein complexes and protein unbound small RNA species was sequenced with NGS. These data provide the first comprehensive tissue-specific miRNA and siRNA expression landscape for a developing fleshy fruit. We identified and several novel, abundantly expressing tissue- and pepper-specific small regulatory RNA species. Our data show that fruit expansion is under extensive tissue-specific sRNA-mediated regulation. However, these experiments are currently still in progress. We completed one manuscript for submission regarding the tissue specific accumulation of sRNA species from total RNA samples. In the close future we intend to publish our different data sets especially focusing on the gel filtration experiments. Moreover, we also investigated technological approaches helping to analyse the biological role of candidate RNA species such as, virus induced gene silencing, artificial miRNA technology and genetic transformation of pepper. We also took part in experiments discovering the sRNA profiles of another *Solenaceae* plant (Nicotiana benthamia), which data can be used as cross reference in further experiments.

References

Johnson NR, Yeoh JM, Coruh C, Axtell MJ. 2016. Improved Placement of Multi-mapping Small RNAs. G3: Genes|Genomes|Genetics **6**, 2103–2111.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology **15**, 550.

Pimentel H, Bray NL, Puente S, Melsted P, Pachter L. 2017. Differential analysis of RNA-seq incorporating quantification uncertainty. Nature Methods **14**, 687–690.

Expansion of *Capsicum annum* fruit is linked to dynamic tissue-specific differential expression of miRNA and siRNA profiles

Dénes Taller[1,†], Jeannette Bálint[1,†], Péter Gyula[1], Tibor Nagy[2], Endre Barta[1], Ivett Baksa[1], György Szittya[1], János Taller[5] and Zoltán Havelda[1,*]

[1] National Agricultural Research and Innovation Centre, Agricultural Biotechnology Institute, Szent-Györgyi Albert út 4., Gödöllő, 2100, Hungary

[2] Department of Plant Science and Biotechnology, Georgikon Faculty, University of Pannonia, Festetics str. 7., H-8360 Keszthely, Hungary

[†] These authors contributed equally to this work.

[*] Correspondence: havelda.zoltan@abc.naik.hu

Dénes Taller, taller.denes@gmail.com

Bálint Jeannette, balint.jeannette@abc.naik.hu

Gyula Péter, gyula.peter@abc.naik.hu

Barta Endre, barta.endre@abc.naik.hu

Baksa Ivett, baksa.ivett@abc.naik.hu

Szittya György, szittya.gyorgy@abc.naik.hu

János Taller, taller@georgikon.hu

Abstract

Background: Small regulatory RNAs, such as microRNAs (miRNAs) and small interfering RNAs (siRNAs) have emerged as important transcriptional and post-transcriptional regulators controlling a wide variety of physiological processes including fruit development. Data are, however limited for their potential roles in developmental processes determining economically important traits of crops. The aim of the current study was to discover and characterise differentially expressed miRNAs and siRNAs in sweet pepper (*Capsicum annuum*) during fruit expansion.

Results: High-throughput sequencing was employed to determine the small regulatory RNA expression profiles in various fruit tissues, such as placenta, seed and flesh at 28 and 40 days after anthesis. Comparative differential expression analyses of conserved, already described and our newly predicted pepper-specific miRNAs revealed that fruit expansion is accompanied by an increasing level of miRNA-mediated regulation of gene expression. Accordingly, ARGONAUTE1 protein, the main executor of miRNA-mediated regulation, continuously accumulated to an extremely high level in the flesh. We also identified numerous pepper-specific, heterochromatin-associated 24-nt siRNAs (het-siRNAs) and hairpin-associated RNAs (ohpRNAs) which were extremely abundant in the seeds, as well as 21-nt and 24-nt phased siRNAs (pha-siRNAs) that were expressed mainly in the placenta and the seeds.

Conclusion: To our knowledge, this work provides the first comprehensive tissue-specific miRNA and siRNA expression landscape for a developing fleshy fruit. We identified several novel, abundantly expressing tissue- and pepper-specific small regulatory RNA species. Our data show that fruit expansion is under extensive tissue-specific sRNA-mediated regulation,

raising the possibility that manipulation of sRNA pathways may be employed to improve the quality and quantity of the pepper fruit.

Keywords: miRNA, phasiRNA, siRNA, *Capsicum annuum*, fruit, development.

Background

The fruit, this highly specialised complex organ, is the presumed crucial factor of the reproduction and dispersal success of angiosperm plants during evolution. Various types of tissues compose the complex structure of the fruit requiring sophisticated, multi-layered developmental processes during fruit formation. Understanding the molecular bases of fruit development processes is very important from a practical point of view as well since different types of fruits provide essential nutrients and minerals for healthy and balanced diet. The discovery that eukaryotic organisms, including plants, employ several RNA interference (RNAi) pathways in which small regulatory (s)RNAs (20 to 24 nucleotide in length) mediate the negative regulation of their target RNAs in a sequence-specific manner broadened our understanding of gene regulation. In plants, RNAi pathways play indispensable roles in developmental processes and in biotic and abiotic stress responses. Distinct RNAi pathways generate different classes of small regulatory RNAs such as microRNAs (miRNAs) and small interfering RNAs (siRNAs) [1-3]. MiRNAs have been revealed as a crucial component in fine regulation of gene expression during developmental processes. MiRNAs modulate gene expression by regulating the degradation or translational repression of their target mRNAs [4]. Fully functional mature miRNAs are generated from their precursors by some phylogenetically conserved protein families [2] in the following manner: the nuclear genome-

coded miRNA genes (*MIR*) are transcribed by RNA polymerase II (Pol II) to produce primary miRNAs (pri-miRNAs) possessing a specific and stable hairpin-like secondary stem-loop structure. In the nucleus, an RNAse III enzyme, DICER-LIKE1 (DCL1) cleaves the hairpin out from the rest of the pri-miRNA resulting in the production of 80–300-nt-long miRNA precursors (pre-miRNAs). In plants, the same DCL1 enzyme then cleaves the precursor at sites determined by structural features to generate small double-stranded miRNA intermediates, the miRNA:miRNA* duplexes. These duplexes are 2'-O-methylated at their 3' termini by the HUA ENHANCER 1 (HEN1) methyltransferase and then transported to the cytoplasm where they are loaded into the RNA-induced silencing complex (RISC). This complex displaces one of the strands (miRNA*) while the mature strand (miRNA) remains bound. The miRNA-containing RISC mediates sequence-specific repression via mRNA-cleavage or translational repression. The main executor of RISCs are the ARGONAUTE (AGO) proteins [5]. The *Arabidopsis thaliana* genome encodes ten AGO proteins which are specialised for various RNAi pathways but often display functional redundancies. AGO1 is the central component of the RISC playing dominant role in miRNA-mediated regulation and its expression is regulated by a feedback mechanism-mediated by miR168 [6]. Beside miRNAs there is a less defined group of hairpin-associated RNAs that cannot be classified as miRNAs (other hairpin RNA, ohpRNA; [7]). This class of sRNAs are numerous, and are mainly derived from inverted repeat elements that are transcribed into long RNA (retrotransposons). They are characterised by longer precursors, larger, more complex loops and less precise maturation than miRNAs. However, they resemble miRNAs in their stranded production (the self-complementary single-stranded precursor folds back to form double-stranded stretches that are recognised by various Dicer-like endonucleases) and that usually there are one or two major RNA species which are much more abundant than the other sequences produced from the same precursor. From ohpRNA loci predominantly 24-nt

sRNAs are produced. The function of this class of sRNA is not known, it is presumed that similarly to heterochromatic siRNAs (hetsiRNAs) they are involved in the repression of transposable elements. The generation of siRNAs is different from that of miRNAs because it is dependent on an RNA-dependent RNA polymerase (RDRP) [3]. The perfect double-stranded (ds)RNA intermediates generated by specific RDRPs can be recognised and cleaved by Dicer-like enzymes other than DCL1 to produce various classes of siRNAs [3]. HetsiRNAs for example are mainly 24-nt in length and generated by DCL3 via processing of Pol IV/RDR2-dependent dsRNA species originating from intergenic or repetitive regions of the genome. These 24-nt-long siRNAs are loaded into the RNA-induced transcriptional silencing (RITS) complex, which is functionally different from the RISC. They recognise Pol V-dependent short transcripts emerging from repetitive and intergenic regions and recruit chromatin-modifying complexes. HetsiRNAs play fundamental roles in maintaining genome integrity, by initiating *de novo* DNA methylation on transposable and repetitive elements mainly during reproductive transitions (meiosis, gametogenesis and embryogenesis), especially in plants with larger, repetitive genomes. Phased (pha)siRNAs, however, are derived from Pol II-dependent mRNAs which are cleaved by a special class of miRNAs/siRNAs of 22-nt length [8]. This cleavage renders one of the cleavage products to be a substrate of RDR6/SUPPRESSOR OF GENE SILENCING 3 (SGS3). The resulting dsRNA is then cleaved by DCL4 endonuclease following a distinctive, regular (phased) pattern starting from the cleavage site. The phasiRNAs are preferentially loaded into a distinctive RISC containing AGO5. A subset of phasiRNAs, the trans-acting siRNAs (tasiRNAs) act mainly at the post-transcriptional level by negatively regulating their target transcripts including gene families which encode disease resistance proteins or transcription factors [9].

The continuously increasing number of experimental studies are demonstrating that plant tissue and organ differentiation is regulated by specific sRNAs. Scientific experiments

proved the importance of miRNAs and tasiRNAs in many aspects of plant development, including leaf, shoot, flower and root development, phase-change, auxin signalling and response to environmental/biotic stresses [10-13]. In addition to the well-known evolutionarily conserved miRNA families, it is extremely important to identify and characterise new, non-conserved miRNAs that are present in closely related species (lineage-specific miRNAs; [7, 14]. The identification of new miRNAs and the genome-wide, tissue-specific analyses of miRNA expression can help to specify miRNA-regulated developmental processes substantially affecting the quality of economically important traits of crop plants [15]. The role of miRNAs in fruit development has been extensively investigated by NGS technology [16-18]. Several studies analysed the expression of miRNAs at a genome-wide level during fruit development in tomato, a *Solanaceae* model plant [19-23]. Fruit development in another important *Solanaceae* species, the pepper (*Capsicum annuum*), is less extensively investigated. Conserved and potentially new, species-specific miRNAs have been described as a result of an investigation of different pepper organs including the developing fruit as a whole [24-27]. However, since the activity of miRNAs is spatiotemporally highly regulated (Nag and Jack, 2010; Garcia, 2008; Valoczi *et al.*, 2006) it is inevitable to get detailed tissue-specific information about the expression pattern of miRNAs during fruit development. Moreover, little is known about the contribution of sRNA-mediated regulation to the fleshy fruit development at the phase of cell expansion where the economically valuable fruit volume is often to reach a two magnitude fold increase.

To get a high-resolution spatiotemporal information about dynamic sRNA expression pattern associated with pepper fruit development in the expansion phase fruit was dissected to get flesh, seed and placenta samples at two time points, 28 and 40 days after anthesis (DAA), and investigated by NGS methodology. Using the gained datasets, we generated the comparative tissue-specific differential expression profiles of the conserved, already described and our

144 newly predicted pepper-specific miRNAs. We found that differential expression activity of
145 miRNAs are intriguingly high at 40 DAA and especially the expansion of flesh is associated
146 with profound changes in the miRNA expression profile. The pronounced role of miRNA-
147 mediated regulation in flesh development was also reflected by the high level of AGO1
148 protein present in flesh samples. Moreover, the analyses of non-miRNA-like sRNAs revealed
149 an ample of pepper-specific hetsiRNAs and phasiRNAs playing a role mainly in the
150 development of the seed and the placenta. As a summary, our work demonstrates that the cell
151 expansion phase of the economically important sweet pepper is under intensive sRNA-
152 mediated regulation implying that breeding technologies capable of modifying the expression
153 of key sRNA regulators can be potent tools to enhance the economic value of pepper.

154 **Methods**

155 **Plant material**

156 „Fehérözön" a Hungarian sweet cultivar of *Capsicum annuum* was used for the experiments.
157 Seeds were germinated on plate between wet filter papers in a growth chamber (stable 21 °C,
158 in darkness). Seedlings were planted into pots and grown in a light room under long day (16 h
159 light/8 h dark) conditions. The day of the anthesis of the selected flowers was recorded and
160 the fruits were collected on the appropriate day after anthesis (DAA). We worked with 2
161 different developmental stages (DAA: 28 ,40).

162 **RNA extraction and small RNA library construction**

163 Total RNA was isolated with using Trizolate Reagent (UD Genomed Ltd., Debrecen,
164 Hungary) from different stages and different tissues. We extracted the samples from fruits, so
165 we could separate with a blade the flesh, the placenta and the seeds. The RNA concentration
166 and 260/280 ratio were determined by using NanoDrop spectrophotometer (Thermo

Scientific, Wilmington, DE) and RNA integrity was verified by agarose gel electrophoresis. 30 µg of total RNA per sample was separated on an 8% denaturing polyacrylamide/urea gel along with RNA size markers and then the sRNA range was isolated. The pellets were dissolved in 20 µl RNase-free ultra pure water, of which 2.5 µl was used for library preparation. The sRNA libraries were constructed using the Truseq Small RNA Sample Prep Kit (Illumina, CA, US), according to the manufacturer's instructions. The pools of eight PCR-amplified, gel-purified and bar-coded cDNAs were submitted to UDGenomed Ltd. (Debrecen, Hungary) for sequencing on Illumina HiSeq 2000 platform.

Bioinformatic analysis

Processing of the raw sequences was carried out using cutadapt 1.9.1. A quality check was performed before and after processing with FastQC 0.11.5 and custom Linux shell scripts. For miRNA analysis, clean reads were aligned to the reference genome of *Capsicum annuum* cv. CM334 (version 1.55). MirCat [28] and miRDeep-P [29] were used in parallel to find new miRNAs, then the two datasets were combined and filtered by custom Phyton scripts. Three subsets were used: miRNAs from the papers about pepper genome and miRNA sequencing, miRNAs from MirProf results and miRNAs from mirCat and miRDeep-P predictions. Raw expression values were determined by Patman alignment and custom Python scripts were used to create a raw abundance and also a normalised (with library size) matrix (Additional Table S1-S3). For differential expression analysis DESeq2 [30] was used with raw abundance matrix. We got the basemean values from DESeq2. We considered those miRNAs as differentially expressed, where the adjusted p-value was less than 0.05. For the visualisation of the results of differential expression analysis, we created MA-plots in Microsoft Excel. The clustering of miRNAs into families was carried out using cd-hit EST by setting the sequence identity cut-off to 0.8. For multiple alignments of the sequences of miRNA families we used Clustal Omega, Color Align Conservation [31] and WebLogo [32] applications. The sRNA

producing genomic clusters were identified using ShortStack 3.4 [33]. This version of ShortStack follows a novel mapping strategy that accurately places multi-mappable reads which is crucial for repetitive genomes. During the alignment we allowed one mismatch and a maximum of 1000 possible placements for multi-mappable reads. In order to allow reporting the predominant RNA size in all clusters the parameters called dicermin and dicermax were set to 18 and 35, respectively. All the other parameters were kept at the default value. The raw read counts in the libraries were rescaled to a library size of one million reads. Global mapping statistics were calculated using Samtools 1.3 and custom Linux shell scripts. Heatmaps were prepared with the R package heatmap.2 and viridis. Genome browser tracks were created using bedtools genomecov command.

Western blot analysis

Tissue samples were homogenised in an ice cold mortar in 400 µl 2× Laemli buffer. Protein samples were resolved on 8% SDS-polyacrylamide gel and blotted to PVDF Blotting Membrane (Amersham™ Hybond™) with Trans-Blot Turbo (BIO-RAD) and subjected to Western blot analysis. The membrane was blocked using 5% non-fat milk powder in PBS containing Tween 20 (PBST) for 1 hour. The target proteins were detected using anti-*Arabidopsis thaliana* AGO-1 (Agrisera). We hybridised the membrane with the antibodies in PBST containing non-fat milk powder at room temperature for 2-3 h. After washing the membrane in PBST, secondary peroxidase-conjugated goat anti-rabbit IgG (H&L) was added. The signals were visualised by chemiluminescence (Clarity™ Western ECL substrate; BIO-RAD) according to the manufacturer's instructions. The membrane was stained with Ponceau reagent to check even protein loading.

Small RNA Northern blot analysis

For the sRNA analysis, 5 μg of total RNA was separated on denaturing 12% polyacrylamide gel containing 8 M urea and transferred to Hybond^{TM}-NX membrane (Amersham) with Fast Blot (Biometra). RNA was chemically crosslinked to the membrane (2008 Pall and Hamilton). We hybridised the membranes radioactively labelled LNA or DNA oligonucleotide probes to detect sRNAs [34]. In all, 10 pmol of each oligonucleotide probe was end-labelled with [γ-^{32}P]ATP by using T4 polynucleotide kinase. The hybridisation was carried out at 37°C (DNA) or 50°C (LNA) depending on the type of oligo nucleotides. Washing of the membrane was done for 5 min two times with washing solution containing 0.1% SDS and 2× SSC at the temperature of the hybridisation. The used oligonucleotide probes are listed in Additional Table S44.

Results and Discussion

Small RNA sequencing and data pre-processing

In this study, we used high-throughput sequencing method to analyse the tissue and development-specific expression of sRNAs and their possible roles in the fruit development of sweet pepper (*Capsicum annuum* cv. Fehérözön). We collected samples at later developmental stages, namely at 28 and 40 days after anthesis (DAA) (Figure 1A). At these stages, the phase of rapid cell division is completed and the development is already in the cell expansion phase. Since cell expansion accounts for the largest increase in fruit volume, the regulation of these developmental stages are extremely important from an economical point of view. To obtain tissue-specific sRNA expression profiles we dissected the pepper fruit. Seed, placenta and flesh tissues were collected separately (Figure 1B and Additional Figure S1) and processed for sRNA sequencing in two independent biological repeats. A total of more than 135 million reads were generated, in the range of 4.4–17.5 million reads per individual

samples (Table 1). After removing adaptor sequences and filtering out low-quality tags, a total of 131.19 million clean reads were obtained, which were 16–28-nt in length. Next, we removed low abundance (<2) reads, and those that aligned to known non-coding RNAs (rRNA, snRNA, snoRNA, tRNA), which resulted in 74.8 million filtered reads. More than 85% of these filtered reads matched the genome (version 1.55) of pepper cultivar CM334 (Qin *et al.*, 2014b) perfectly. The reasons for presence of non-matching reads can be that the reference genome is only closely related and incomplete (i.e. mitochondrial and chloroplastic reference sequences were masked from the finished genome), there might be Fehérözön-specific sequences that are missing from CM334 and that antiviral siRNAs (mostly 21-22-nt-long) derived from a resident Bell pepper endornavirus also contributes to the siRNA pool. The most critical filtering step was the removal of sequences with low abundance. The small numbers of sequences with an invalid length, low-quality tags, or low complexity indicate that the quality of our libraries were appropriate. Jaccard-similarity matrix generated from the 5,000 most abundant sequences revealed that the biological replicates clustered together properly, except in the case of placenta samples (Additional Figure S2). However, this can be explained by the particular sRNA expression profile characteristic for the placenta (see later).

Size distribution of small RNA reads

The size distribution of all the filtered reads also confirms the high-quality of sRNA sequencing data since more than the 94% of the filtered reads falls into the 21–24-nt-long category and only 5.82% of the reads represents the 16–20 (cca. 4.8%) and 25–28-nt-long (cca 1%) reads (Figure 2). The 46.73% of the total sRNAs were 24-nt-long while 24.05% were 21-nt-long. The 22 and 23-nt-long sRNAs were the less abundant classes, 14% and 10%, respectively (Figure 2). In accordance with previously reported studies, the analyses of the individual samples revealed that the 24-nt-long sRNA species accumulated most dominantly in the majority of the samples [7, 35]. However, as an exception, we observed the dominant

accumulation of 21-nt-long sRNAs in the flesh at 28 DAA and even more prominently at 40 DAA (Figure 2) indicating that the tissue-specific relative proportion of different size classes of sRNAs can vary greatly during the pepper fruit development. To analyse this phenomenon, we calculated the 21/24 ratio of the different samples and observed differences between various tissue types and/or developmental stages. The comparative 21/24 ratios of samples deriving from placenta at 28 and 40 DAA showed only a moderate increase. However, in contrast to this relatively constant 21/24 ratio of placenta the 21/24 ratio decreased in developing seeds from 55% (seed 28 DAA) to about 25% (seed 40 DAA) while the opposite change was observed in the flesh where the ratio increased from 75% (flesh 28 DAA) to about 200% (flesh 40 DAA). In general, the rate of the 21-nt-long sRNAs is increasing along the expansion of the fruit with the exception of the seed. The dynamic changes in the relative accumulation of various sRNA species with the advance of time indicate that sRNAs may play a pivotal role during fruit expansion. Moreover, the high percentage of the 21-nt-long sRNA species in the flesh suggests that miRNAs, phasiRNAs or other 21-nt-long sRNA molecules play a crucial role in the development of the economically most important tissue of the pepper fruit.

Conserved miRNAs in the developing pepper fruit

Two approaches have been applied to identify known miRNAs in the pepper fruit-specific sRNA libraries. MirProf has been used to profile known miRNAs by allowing maximum two mismatches. MirProf uses Patman to align reads against miRBase [36]. MiRBase currently does not contain any pepper miRNAs, so we aligned our reads to the recently published conserved pepper miRNAs (can-mir) using Patman with the same parameters [24, 26]. We discarded sRNA reads, which were not present at least in two libraries (213 different sequences remained), however this dataset contains some 18 nt long sequences, which were set aside. Our final set of conserved miRNAs contained 193 non-redundant sequences, from

38 conserved miRNA families (Additional Table S1). MirBase contains three species (*Solanum tuberosum, Solanum lycopersicum, Nicotiana tabacum)* from the *Solanaceae* family. The majority of these miRNAs can be found in mirBase with two mismatches as miRNAs described in three *Solanaceae* species (*Solanum tuberosum (143), Solanum lycopersicum* (132), *Nicotiana tabacum (110)).* A recent comprehensive survey of plant miRNAs from 31 vascular plants species specified 82 known miRNA families forming eight groups depending on their distribution across lineages and species [37]. The first group contains those miRNA families which are widely present and highly expressed in all terrestrial species. In our libraries we found that all of these miRNA families showed high level expression profiles, exhibiting more than 75 reads per million (RPM) at least in one library (sum of all family members) (miR156 (249 RPM), miR166 (466210 RPM), miR167 (95 RPM), miR168 (4444 RPM), only the miR172 family showed a lower expression level (17 RPM). The group 2 miRNA families are represented in all taxonomic lineages, but can be absent or present with low abundance in some species (miR158, miR159, miR160, miR162, miR164, miR169, miR171, miR319, miR390, miR393, miR394, miR396, miR397, miR529, miR535 and miR4414). Most of the group 2 miRNA families were also present in our data sets with the exception of the miR158, miR529 and miR535 families. Seven families (miR159, miR162, miR169, miR171, miR319, miR393, miR396) were abundant (more than 75 RPM) at least in one library. Other six families (miR160, miR164, miR390, miR394, miR397, miR4414) were present with low abundance (around 5-50 RPM). MiR158, miR529 and miR535 were absent from all the investigated pepper libraries. In a recent study investigating *Nicotiana benthamiana* sRNAs, it was also shown that miR529 together with miR535 were absent from sRNA libraries [38]. One of the other six groups, group 8, was enriched in *Solanaceae* species, as *Petunia*, *Nicotiana tabacum*, *Solanum lycopersicum*, *Solanum tuberosum* and *Capsicum annuum*. From the 10 miRNA families present in this

group (miR1919, miR4376, miR5300, miR5301, miR6022, miR6023, miR6024, miR6025, miR6026, miR6027, miR6149), we identified 5 (miR4376, miR5300, miR5301, miR6024, miR6027) in our libraries.

Known pepper-specific miRNAs in the developing pepper fruit

Recent studies described new pepper-specific miRNAs [24, 26, 27]. Based on published pepper miRNA sequences, we prepared a merged list of known pepper-specific miRNA sequences. The latest paper on pepper fruit miRNAs [27] described four sequences (Novel_miR265, Novel_miR218, Novel_miR23, Novel_miR279) from our set of new miRNAs, so we transferred these miRNAs to the known pepper-specific group, with the ID given in that paper. The other published two sets of new miRNA sequences have only two common elements present in their newly described miRNAs, (can-miR-n010/can-miRC7-3p.1 [UUAUGAGAUAAGUUCAACACG], can-miR-n003a/can-miRC13-5p.1 [UGUAGUUGUAGCCAUUCUAUU]). A non-redundant list of 68 sequences was formed and 50 of these sequences (the common two are involved) were found in our pepper fruit libraries (Additional Table S2). We found more than 83% of the miRNAs described by Hwang et al. and 33% of the miRNAs described by Qin et al.[24, 26]. Some of the miRNAs, such as can-mir-n05 and can-mir-n027, were not found in our libraries. This can be explained by the expression patterns of these miRNAs since they show higher abundance in root and stem libraries. Nonetheless, we detected known new miRNAs which have displayed strong flower/leaf-specificity or low expression in previously investigated fruit samples (for example, can-mir-n019a-d or can-mir-n030).

New miRNAs in developing pepper fruit

To reveal still unknown miRNAs, we constructed a novel miRNA identification pipeline (Additional Figure S3). 74.82 million filtered reads were used to identify new miRNAs by

paralleled utilisation of two computational plant miRNA prediction tools, miRDeep-P and miRCat. Both of these tools map the reads to the genome and extract reference sequences by extending mapped reads in the flanking regions based on their ability to fold predicted miRNA precursor RNA stem-loop structures. These programs use the Vienna RNA package [39] to determine the secondary structure of the predicted precursors and apply a scoring system to rank the obtained potential miRNAs precursor structures. Although these programs use similar computational approaches they can generate significantly different results. To enhance the reliability of our miRNA predictions, we used both programs to identify the potentially new pepper miRNAs and accepted only the overlapping set of different predictions. The next filtering steps were the elimination of candidate reads which did not reach the abundance level of 10 RPM and appeared at least in two sRNA libraries. Utilisation of this method allowed us to describe 73 different, highly confident novel pepper miRNA sequences (Additional Table S3). All of these novel miRNAs possess predicted miRNA* strands, which were found in at least one of our sequenced libraries, a stable secondary structure of the predicted precursors (Additional Table S4). The majority of the newly identified miRNAs can not be assigned to families. The rest of the sequences were clustered into 6 miRNA families (Additional Figure S4). We also checked the similarity between our newly described and the previously described miRNAs, and found that five (can-mir-f11, can-mir-f14, can-mir-f27, can-mir-f46, can-mir-f57) of our miRNAs were similar to miRNAs described previously by [27]. These family members differ mainly in their sequence composition indicating that they are deriving from different precursors. The structures of some selected novel miRNA representatives are shown in Figure 3. The emergence of novel miRNA families may represent an important evolutionary step in the fixation of miRNA-mediated control mechanism.

Validation of miRNA expression profiles

sRNA Northern blot analyses were carried out to validate experimentally the tissue-specific and developmental-stage-specific expression patterns of miRNAs determined by the bioinformatics analyses of high-throughput sequencing data. Total RNA samples deriving from seed, placenta and flesh at 28 and 40 DAA were used for sRNA Northern blot experiments. DNA or LNA oligonucleotide probes were used to detect the expression patterns of selected conserved miRNAs (miR171, miR172, miR396, miR159 and miR167) (Figure 4A). The signal intensity of Northern analysis results revealed a good agreement with the sequencing data. In some cases, we observed discrepancies between the Northern blot and sequencing data. These discrepancies could be attributed to the fact that the hybridisation-based technology measures the abundance of heterogeneous miRNA population depending on the parameters of the hybridisation procedures. Next we analysed the expression patterns of selected new pepper-specific miRNAs (Figure 4B). In line with the conserved miRNA expression data the experimental spatiotemporal analyses of these novel miRNAs revealed highly tissue specific expression patterns in accordance with tendency of the sRNA sequencing data. Altogether these data indicate that the gained sRNA sequencing data are suitable for differential expression analysis of miRNAs during the developmental process.

Differential miRNA expression in the pepper fruit

The high-throughput sequencing allows the use of RPM data to define the miRNA expression levels in the investigated tissues and developmental stages. To gain insights in the putative roles of the identified miRNAs the relative differential expression patterns of miRNAs were investigated by comparing different tissue types and developmental stages. We investigated the differential expression of conserved miRNAs, known pepper-specific miRNAs and our newly identified pepper-specific miRNAs using DESeq2 MA-plots to reveal the changes in

the miRNA expression levels at the investigated time points. First we compared the relative expression of various miRNA categories at 28 DAA (Figure 5, 6 and 7, panel A). In the placenta very few changes were detected relative to flesh and only moderate changes were observed relative to seed (Additional Table S5-S6, S14-S15, S23-S24). In these cases we detected the altered expression levels of only few potentially influential conserved miRNAs. No new or pepper-specific miRNAs showed differential expression in the placenta relative to the seed and these classes of miRNAs displayed moderate expression changes in the placenta relative to the seed. More pronounced changes were detected between the flesh and seed samples where 10 conserved miRNAs, several new and known pepper-specific miRNAs displayed significantly altered expression (Additional Table S7, S16, S25). Altogether at this time point we did not observe extensive differences in the expression of the investigated miRNAs. The most extensive changes were observed between the flesh and the seed indicating that miRNA driven regulation started to diverge in these tissues. However, at 40 DAA, we observed extensive changes in expression profiles of all miRNA classes between every tissue (Figure 5, 6, 7, panel B, Additional Table S8-10, S17-19, S26-28). This suggests that miRNA-mediated regulation processes are increasingly active during the expansion phase. To reveal the contribution of different tissue types to the relative miRNA expression changes, we also compared the miRNA expression profiles of the identical tissue types at the two investigated time points, 28 and 40 DAA (Figure 5, 6 and 7, panel C). We observed significant differences in the case of the seed and the flesh samples (Additional Table S12-13, S21-22, S30-31). The placenta, however, showed practically no significant miRNA expression changes (Additional Table S11, S20, S29). The fact that the flesh samples showed similar changes in the miRNA expression profiles in every miRNA classes indicates that the economically most important tissue type of the pepper fruit is under an extensive control of miRNA-mediated regulation during the expansion phase.

Expression of AGO1 protein in pepper fruit tissues

We found that pepper fruit flesh expansion is associated with profound temporal changes in the miRNA expression profile which can indicate the central role of miRNA-mediated regulation in this developmental process. AGO1 is the most important AGO protein in the miRNA pathway which is responsible for the cleavage or translational inhibition of target mRNAs determined by the loaded miRNA. AGO1 homeostasis is controlled by the action of miR168 on *AGO1* mRNA [6]. Next we investigated the accumulation of AGO1 protein, as the key executor component of RNAi, postulating that the accumulation level of AGO1 reflects the activity RNAi in the particular tissue types. First we analysed the accumulation of the AGO1 key regulator miR168 by sRNA Northern blot. We found, in agreement of the NGS data, that miR168 expresses abundantly in general with the highest level in the flesh samples at both time points (Figure 8). Using protein extracts representing the same samples we analysed the expression of AGO1 with western blot. At 28 DAA no significant AGO1 accumulation was detected in the seed and the placenta samples at the detection level of the used method while in the flesh sample AGO1 accumulated at very high level. At 40 DAA we experienced drastic and moderate AGO1 induction in the seed and the placenta samples and importantly a continuously high AGO1 expression in the flesh. In accordance with the differential expression results, these data indicate that miRNA-mediated regulation plays a pivotal role in the expansion of the pericarp. The importance of miRNAs in expansion of the flesh is further supported by the observation that the proportions of the 21-nt-long miRNA-like molecules are highest in the flesh NGS data, especially at 40 DAA, where the 21/24-ratio becomes higher than one (Figure 2). In the placenta, the AGO1 level remains relatively low suggesting that temporal changes in the miRNA expressions are not required for the proper formation of this tissue. Indeed, the differential expression of the pattern of the placenta samples at the two investigated time points showed no changes. In the case of the seed, very

low AGO1 expression is detected at the early time point, however in the later time point the amount of the AGO1 protein drastically increased probably due to the increasing size of the embryo (Supplemtary Figure S1). We also found that the levels of miR168 in the samples do not correlate with the amount of AGO1 protein suggesting a potentially unusual regulation. This phenomenon requires further investigation.

Expression of 24-nt-long hc-siRNA and ohpRNA clusters

The most abundant class of endogenous siRNAs are 24-nt-long and are predominantly produced from transposons and repeated sequence elements, especially heterochromatic regions [40]. Genome associated siRNAs are engaged in various nuclear processes such as developmental gene and transposon regulation, heterochromatin formation, and genome stability [41]. It has also been shown that 24-nt-long siRNAs play a crucial role in transgenerational epigenetic inheritance [42]. To analyse the expression pattern of the 24-nt siRNA-producing clusters during pepper fruit expansion we investigated our datasets using ShortStack [33]. We identified 1,012,092 sRNA-producing clusters altogether. Most of the *de novo* identified siRNA clusters associated with the dominant sequence length of 24-nt (850,801, 84.1%). Of these clusters, 426,309 (42.1%) appeared to be associated with sequence repeats or mobile genetic elements. Others are localised to the promoters or gene bodies of protein coding genes, probably affecting the methylation state of the DNA and consequently the transcriptional activity. We performed a differential expression analysis on the non-miRNA producing sRNA clusters using DESeq2 in the same way as in the case of miRNAs. Pairwise expression changes are graphically represented on MA-plots to visualise the distribution of expression values and differences (Figure 9). The associated information provided by ShortStack and DESeq2 for the significantly changed clusters are listed in the Additional Table S32-40. We can observe a similar trend in changes between tissues and time

points as in the case of miRNAs. Most importantly, there are only a small number of significantly changed clusters in the placenta samples between the two time points, a moderate number of changing clusters between the placenta and the flesh samples, whereas the greatest changes can be observed between the seed and other tissues. Also it is noteworthy that the expression levels of sRNA clusters are generally higher in the seed samples at 40 DAA than at 28 DAA. This can be explained by the growth and differentiation of the embryonic tissue in the seed (Additional Figure S1), suggesting that the 24-nt-long hc-siRNAs or ohpRNA-producing loci are more active in this tissue in order to prevent genome instability during embryonic development.

The detailed analyses of the top 50 most abundant 24-nt hc-siRNA/ohpRNA-producing clusters revealed tissue-specific and temporally highly regulated expression patterns (Figure 10A, Additional Table S41). Some siRNA/ohpRNA-producing loci expressed at a very high level in the seed at 28 DAA, drastically reduced at 40 DAA, and almost completely silent in other tissues. To verify these remarkable tissue- and developmental stage-specific expression of the named clusters we investigated the expression of two representative clusters by sRNA Northern blot analyses (Figure 9C). Cluster 297979 is pepper-specific, covering approximately 5 Kb-long intergenic region and expresses at a very high level. The cluster is stranded but cannot be folded into a simple hairpin structure. A dominant 24-nt-long siRNA species is produced from the negative strand. Cluster 461037, approximately 1 Kb long, exhibits an opposite pattern of expression compared to Cluster 297979, showing enhanced expression at 40 DAA in the seed (Figure 9C). This locus was described earlier as a miRNA gene (can-miR-n008) [24] but in our dataset it looks more like an ohpRNA cluster. The dominant 24-nt-long sequence contains one mismatch compared to the can-miR-n008 mature sequence. Furthermore, no star sequence was found in any libraries. Under our experimental conditions, an abundant 21-nt-long siRNA species is also produced. In line with the NGS

485 data, we detected the enhanced expression of both abundant sequence species in the seed at 28
486 DAA which declined at 40 DAA using DNA probe specific to the dominant siRNA species
487 (Figure 9C). At this sensitivity level, in agreement with the NGS data, both probes detected
488 no signals in the flesh and the placenta samples.
489 Altogether, these data reveal a dynamic spatiotemporal regulation of the expression of
490 genome associated siRNAs in the developing pepper fruit. These observations are in
491 accordance with previous Arabidopsis results demonstrating the dynamic expression of Pol
492 IV-dependent siRNAs during silique development [43]. It has been shown that the Pol IV-
493 dependent siRNA accumulation is initiated in the maternal gametophyte and continues in the
494 seed, predominantly in the developing endosperm. Our observation that siRNA clusters are
495 activated in a developmental phase-dependent and tissue-specific manner suggests that
496 genome-associated siRNAs are potent regulatory components of the pepper fruit
497 development.
498
499 **Expression of phased siRNA and secondary siRNA-producing clusters**
500
501 Phased secondary siRNAs are produced from Pol II-dependent transcripts that are cleaved by
502 a predominantly 22-nt-long phase-initiating miRNA [44]. One exception is TAS3, which is
503 cleaved by a 21-nt-long mir390 by a different mechanism [45]. This cleavage then initiates
504 the synthesis of a complementary RNA by an RNA-dependent RNA polymerase (RDR6 in
505 Arabidopsis), resulting in a double-stranded intermediate that is diced by the DCL4
506 endonuclease at every 21-nt starting from the miRNA cleavage site. ShortStack identified 93
507 and 948 potential phased clusters of 21-nt and 24 nt, respectively, the 50 most abundant are
508 shown in Additional Figure S5 and S6 (the associated data are in Additional Table S42-43).
509 The loci were annotated using BLAST comparing the locus sequences to the annotated RNA

set of the Zunla-1 v2.0 genome, because the annotation of CM334 contains only protein coding genes without classification. According to this, the majority of the 21-nt PHAS loci overlap with genes coding for leucin-rich repeat-containing receptors, members of a large resistance protein family. They have already been described as phasiRNA-producing genes in other species. The phasing is usually initialised by miRNAs belonging to the mir482/mir2118 family or by other phasiRNAs. The fact that these mRNAs are diced suggests that the full-length mRNAs are present at low level and therefore these receptors may have no role in fruit development. Surprisingly, we could not identify TAS homologue genes by BLAST alignment of Arabidopsis or other known plant TAS sequences to the Zunla RNA sequences, possibly because only the functional tasiRNA sequences are conserved, not the full noncoding transcripts. We searched for short matching patterns to the known tasiRNA sequences [46] using Patman [47]. This way we managed to identify three possible TAS3 homologue genes (ARF2-like, ARF3, ARF4), two TAS4 homologous genes (Potassium transporter 5 and Aquaporin PIP2-1-like) with microhomology only to the conserved tasiRNA sequences, and two TAS1-like genes (AT4G04775-like uncharacterized protein and ATP-dependent DNA helicase Q-like 5) that can only be found in apple and peach, according to the tasiRNA database [46]. We did not find these potential TAS loci to produce phased siRNAs, probably because their known phase initialising miRNAs are hardly expressed in all of our libraries. Beside the above mentioned large protein family, there are single phasiRNA-producing genes, that are more specific to certain plants. It is worth mentioning that a *DICER-LIKE PROTEIN 2-like* gene is also a phased siRNA producing locus under our experimental conditions, which was also found to be a PHAS locus in soybean and *Medicago truncatula* [9] [48]. Moreover, this locus was differentially expressed between 28 and 40 DAA, in all tissues. DCL2 is a member of the sRNA processing machinery that plays role in endogenous and virus-derived siRNA production. 24-nt PHAS loci have been described earlier [49, 50] and appear to be

developmentally regulated. It is noteworthy that the most abundant 24-nt PHAS loci we identified are also developmental stage-specific, they are much more abundant in the seed sample at 28 DAA (Additional Figure S6). According to the previous findings, these loci are associated with transposons and their function is, similarly to the PIWI-associated siRNA in animals, to maintain genomic stability in reproductive tissues.

Conclusion

In conclusion, to the best of our knowledge, this work provides the first comprehensive tissue-specific miRNA and siRNA expression landscape for developing fleshy fruit. The separated analyses of fruit tissue components allowed us to dissect and list the sRNA components dominantly expressing in different tissue types in the expanding pepper fruit at two time points. Our analyses revealed several abundantly expressing tissue- and pepper-specific small regulatory RNA species implicating their regulatory action. Altogether, the presented data extend the perspectives of sRNA-mediated regulatory processes during the complex program of fruit development and provide fundamental data sets for further experiments.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and analysed during the current study are available in the National Center for Biotechnology Informations (NCBI) BioProject database under Accession number PRJNA388388.

https://www.ncbi.nlm.nih.gov/bioproject/PRJNA388388

https://submit.ncbi.nlm.nih.gov/subs/sra/SUB2736714/overview

This Sequence Read Archive (SRA) submission will be released on 2017-12-31 or upon publication, whichever is first.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by grants from the National Research, Development and Innovation Office (NKFI-K109438, NKFI-K116602 and NKFI-K119701).

Author contributions

D.T. and J.B. prepared the sequencing libraries and conducted the experiments. I.B. and G.S. provided technical support for sequencing library preparation. D.T., T.N. and E.B. performed the miRNA data analysis, P.G. is responsible for the siRNA data analysis. Z.H. designed the project and wrote the manuscript. D.T., J.B., P.G. contributed to the manuscript writing. J.T. helped with data analysis and critically revised the manuscript. All authors read and approved the final version of the manuscript.

Funding

This work was supported by grants from the National Research, Development and Innovation Office (NKFI-K109438, NKFI-K116602 and NKFI-K119701).

Additional files

Figure S1. Close-up view of tissues of the pepper fruit at 28 and 40 DAA.

Figure S2. Jaccard similarity matrix of the small RNA libraries.

Figure S3. Flowchart of the steps of new pepper-specific miRNA prediction.

Figure S4. Alignment and sequence logos of novel miRNA family members.

Figure S5. Expression pattern and sequence length composition of 21-nt-long phased siRNA clusters.

Figure S6. Expression pattern and sequence length composition of 24-nt-long phased siRNA clusters.

Table S1: Conserved miRNA expression in 28DAA and 40 DAA pepper fruit tissues, grouped by miRNA family. The underlined names were given by previous works.

Table S2: Tissue specific expression of previously described pepper miRNAs

Table S3: Expression values of novel pepper miRNAs in pepper fruit tissues

Table S4: Predicted precursor sequences of novel miRNAs

Table S5-S13: Data tables of conserved miRNA differential expression

Table S14-S22: Data tables of known miRNA differential expression

Table S23-S31: Data tables of novel miRNA differential expression

Table S32-S40: Data tables of Small RNA clusters differential expression

Table S41: Top 50 most abundant 24-nt siRNA clusters used for heatmap generation

Table S42: Top 50 most abundant 21-nt phased siRNA clusters used for heatmap generation

Table S43: Top 50 most abundant 24-nt phased siRNA clusters used for heatmap generation

Table S44: List of oligonucleotides used for Northern blot

References

1. Martinez de Alba AE, Elvira-Matelot E, Vaucheret H: Gene silencing in plants: a diversity of pathways. *Biochimica et biophysica acta* 2013, 1829(12):1300-1308.

2. Rogers K, Chen X: Biogenesis, turnover, and mode of action of plant microRNAs. *The Plant cell* 2013, 25(7):2383-2399.

3. Borges F, Martienssen RA: The expanding world of small RNAs in plants. *Nature reviews Molecular cell biology* 2015, 16(12):727-741.

4. Bologna NG, Voinnet O: The diversity, biogenesis, and activities of endogenous silencing small RNAs in Arabidopsis. *Annual review of plant biology* 2014, 65:473-503.

5. Zhang H, Xia R, Meyers BC, Walbot V: Evolution, functions, and mysteries of plant ARGONAUTE proteins. *Current opinion in plant biology* 2015, 27:84-90.

6. Vaucheret H: AGO1 homeostasis involves differential production of 21-nt and 22-nt miR168 species by MIR168a and MIR168b. *PloS one* 2009, 4(7):e6442.

7. Axtell MJ: Classification and comparison of small RNAs from plants. *Annual review of plant biology* 2013, 64:137-159.

8. Chen HM, Chen LT, Patel K, Li YH, Baulcombe DC, Wu SH: 22-Nucleotide RNAs trigger secondary siRNA biogenesis in plants. *Proceedings of the National Academy of Sciences of the United States of America* 2010, 107(34):15269-15274.

9. Fei Q, Xia R, Meyers BC: Phased, secondary, small interfering RNAs in posttranscriptional regulatory networks. *The Plant cell* 2013, 25(7):2400-2415.

10. Dotto MC, Petsch KA, Aukerman MJ, Beatty M, Hammell M, Timmermans MC: Genome-wide analysis of leafbladeless1-regulated and phased small RNAs underscores the importance of the TAS3 ta-siRNA pathway to maize development. *PLoS genetics* 2014, 10(12):e1004826.

11. Kumar R: Role of microRNAs in biotic and abiotic stress responses in crop plants. *Applied biochemistry and biotechnology* 2014, 174(1):93-115.

12. Zhang B: MicroRNA: a new target for improving plant tolerance to abiotic stress. *Journal of experimental botany* 2015, 66(7):1749-1761.

13. Li C, Zhang B: MicroRNAs in Control of Plant Development. *Journal of cellular physiology* 2016, 231(2):303-313.

14. Qin Z, Li C, Mao L, Wu L: Novel insights from non-conserved microRNAs in plants. *Frontiers in plant science* 2014, 5:586.

15. Kamthan A, Chaudhuri A, Kamthan M, Datta A: Small RNAs in plants: recent development and application for crop improvement. *Frontiers in plant science* 2015, 6:208.

16. Xu Q, Liu Y, Zhu A, Wu X, Ye J, Yu K, Guo W, Deng X: Discovery and comparative profiling of microRNAs in a sweet orange red-flesh mutant and its wild type. *BMC genomics* 2010, 11:246.

17. Li H, Mao W, Liu W, Dai H, Liu Y, Ma Y, Zhang Z: Deep sequencing discovery of novel and conserved microRNAs in wild type and a white-flesh mutant strawberry. *Planta* 2013, 238(4):695-713.

18. Ge A, Shangguan L, Zhang X, Dong Q, Han J, Liu H, Wang X, Fang J: Deep sequencing discovery of novel and conserved microRNAs in strawberry (Fragariaxananassa). *Physiologia plantarum* 2013, 148(3):387-396.

19. Mohorianu I, Schwach F, Jing R, Lopez-Gomollon S, Moxon S, Szittya G, Sorefan K, Moulton V, Dalmay T: Profiling of short RNAs during fleshy fruit development reveals stage-specific sRNAome expression patterns. *The Plant journal : for cell and molecular biology* 2011, 67(2):232-246.

20. Moxon S, Jing R, Szittya G, Schwach F, Rusholme Pilcher RL, Moulton V, Dalmay T: Deep sequencing of tomato short RNAs identifies microRNAs targeting genes involved in fruit ripening. *Genome research* 2008, 18(10):1602-1609.

21. Zuo J, Zhu B, Fu D, Zhu Y, Ma Y, Chi L, Ju Z, Wang Y, Zhai B, Luo Y: Sculpting the maturation, softening and ethylene pathway: the influences of microRNAs on tomato fruits. *BMC genomics* 2012, 13:7.

22. Karlova R, van Haarst JC, Maliepaard C, van de Geest H, Bovy AG, Lammers M, Angenent GC, de Maagd RA: Identification of microRNA targets in tomato fruit development using high-throughput sequencing and degradome analysis. *Journal of experimental botany* 2013, 64(7):1863-1878.

23. Gao C, Ju Z, Cao D, Zhai B, Qin G, Zhu H, Fu D, Luo Y, Zhu B: MicroRNA profiling analysis throughout tomato fruit development and ripening reveals potential regulatory role of RIN on microRNAs accumulation. *Plant biotechnology journal* 2015, 13(3):370-382.

24. Hwang DG, Park JH, Lim JY, Kim D, Choi Y, Kim S, Reeves G, Yeom SI, Lee JS, Park M *et al*: The hot pepper (Capsicum annuum) microRNA transcriptome reveals novel and conserved targets: a foundation for understanding MicroRNA functional roles in hot pepper. *PloS one* 2013, 8(5):e64238.

25. Kim S, Park M, Yeom SI, Kim YM, Lee JM, Lee HA, Seo E, Choi J, Cheong K, Kim KT *et al*: Genome sequence of the hot pepper provides insights into the evolution of pungency in Capsicum species. *Nature genetics* 2014, 46(3):270-278.

26. Qin C, Yu C, Shen Y, Fang X, Chen L, Min J, Cheng J, Zhao S, Xu M, Luo Y *et al*: Whole-genome sequencing of cultivated and wild peppers provides insights into Capsicum domestication and specialization. *Proceedings of the National Academy of Sciences of the United States of America* 2014, 111(14):5135-5140.

678 27. Liu Z, Zhang Y, Ou L, Kang L, Liu Y, Lv J, Wei G, Yang B, Yang S, Chen W *et al*:
679 Identification and characterization of novel microRNAs for fruit development and
680 quality in hot pepper (Capsicum annuum L.). *Gene* 2017, 608:66-72.

681 28. Stocks MB, Moxon S, Mapleson D, Woolfenden HC, Mohorianu I, Folkes L,
682 Schwach F, Dalmay T, Moulton V: The UEA sRNA workbench: a suite of tools for
683 analysing and visualizing next generation sequencing microRNA and small RNA
684 datasets. *Bioinformatics* 2012, 28(15):2059-2061.

685 29. Yang X, Li L: miRDeep-P: a computational tool for analyzing the microRNA
686 transcriptome in plants. *Bioinformatics* 2011, 27(18):2614-2615.

687 30. Love MI, Huber W, Anders S: Moderated estimation of fold change and dispersion for
688 RNA-seq data with DESeq2. *Genome Biol* 2014, 15(12):550.

689 31. Stothard P: The sequence manipulation suite: JavaScript programs for analyzing and
690 formatting protein and DNA sequences. *BioTechniques* 2000, 28(6):1102, 1104.

691 32. Crooks GE, Hon G, Chandonia JM, Brenner SE: WebLogo: a sequence logo
692 generator. *Genome research* 2004, 14(6):1188-1190.

693 33. Axtell MJ: ShortStack: comprehensive annotation and quantification of small RNA
694 genes. *Rna* 2013, 19(6):740-751.

695 34. Varallyay E, Burgyan J, Havelda Z: MicroRNA detection by northern blotting using
696 locked nucleic acid probes. *Nature protocols* 2008, 3(2):190-196.

697 35. Parent JS, Martinez de Alba AE, Vaucheret H: The origin and effect of small RNA
698 signaling in plants. *Frontiers in plant science* 2012, 3:179.

699 36. Kozomara A, Griffiths-Jones S: miRBase: annotating high confidence microRNAs
700 using deep sequencing data. *Nucleic acids research* 2014, 42(Database issue):D68-73.

701 37. Chavez Montes RA, de Fatima Rosas-Cardenas F, De Paoli E, Accerbi M, Rymarquis
702 LA, Mahalingam G, Marsch-Martinez N, Meyers BC, Green PJ, de Folter S: Sample
703 sequencing of vascular plants demonstrates widespread conservation and divergence
704 of microRNAs. *Nature communications* 2014, 5:3722.

705 38. Baksa I, Nagy T, Barta E, Havelda Z, Varallyay E, Silhavy D, Burgyan J, Szittya G:
706 Identification of Nicotiana benthamiana microRNAs and their targets using high
707 throughput sequencing and degradome analysis. *BMC genomics* 2015, 16:1025.

708 39. Lorenz R, Bernhart SH, Siederdissen CHZ, Tafer H, Flamm C, Stadler PF, Hofacker
709 IL: ViennaRNA Package 2.0. *Algorithm Mol Biol* 2011, 6.

710 40. Kasschau KD, Fahlgren N, Chapman EJ, Sullivan CM, Cumbie JS, Givan SA,
711 Carrington JC: Genome-wide profiling and analysis of Arabidopsis siRNAs. *PLoS*
712 *biology* 2007, 5(3):e57.

713 41. Castel SE, Martienssen RA: RNA interference in the nucleus: roles for small RNAs in
714 transcription, epigenetics and beyond. *Nature reviews Genetics* 2013, 14(2):100-112.

42. Bond DM, Baulcombe DC: Small RNAs and heritable epigenetic variation in plants. *Trends in cell biology* 2014, 24(2):100-107.

43. Mosher RA, Melnyk CW, Kelly KA, Dunn RM, Studholme DJ, Baulcombe DC: Uniparental expression of PolIV-dependent siRNAs in developing endosperm of Arabidopsis. *Nature* 2009, 460(7252):283-286.

44. Cuperus JT, Carbonell A, Fahlgren N, Garcia-Ruiz H, Burke RT, Takeda A, Sullivan CM, Gilbert SD, Montgomery TA, Carrington JC: Unique functionality of 22-nt miRNAs in triggering RDR6-dependent siRNA biogenesis from target transcripts in Arabidopsis. *Nature structural & molecular biology* 2010, 17(8):997-1003.

45. Montgomery TA, Howell MD, Cuperus JT, Li D, Hansen JE, Alexander AL, Chapman EJ, Fahlgren N, Allen E, Carrington JC: Specificity of ARGONAUTE7-miR390 interaction and dual functionality in TAS3 trans-acting siRNA formation. *Cell* 2008, 133(1):128-141.

46. Zhang C, Li G, Zhu S, Zhang S, Fang J: tasiRNAdb: a database of ta-siRNA regulatory pathways. *Bioinformatics* 2014, 30(7):1045-1046.

47. Prufer K, Stenzel U, Dannemann M, Green RE, Lachmann M, Kelso J: PatMaN: rapid alignment of short sequences to large databases. *Bioinformatics* 2008, 24(13):1530-1531.

48. Fei Q, Li P, Teng C, Meyers BC: Secondary siRNAs from Medicago NB-LRRs modulated via miRNA-target interactions and their abundances. *The Plant journal : for cell and molecular biology* 2015, 83(3):451-465.

49. Song X, Li P, Zhai J, Zhou M, Ma L, Liu B, Jeong DH, Nakano M, Cao S, Liu C *et al*: Roles of DCL4 and DCL3b in rice phased small RNA biogenesis. *The Plant journal : for cell and molecular biology* 2012, 69(3):462-474.

50. Zhai J, Zhang H, Arikit S, Huang K, Nan GL, Walbot V, Meyers BC: Spatiotemporally dynamic, cell-type-dependent premeiotic and meiotic phasiRNAs in maize anthers. *Proceedings of the National Academy of Sciences of the United States of America* 2015, 112(10):3146-3151.

749

750 **Tables**

		Raw	Trimmed	Filter1	Filter2	Aligned
28 DAA	Seed 1	17.51	17.33	11.29	10.72	9.00
	Seed 2	12.18	11.79	8.43	7.90	6.82
	Placenta 1	6.18	5.44	2.78	2.65	2.30
	Placenta 2	16.22	16.08	8.54	8.39	7.39
	Flesh 1	11.69	10.42	9.56	9.03	6.49
	Flesh 2	4.41	4.02	2.42	2.26	1.99
40 DAA	Seed 1	9.27	9.15	4.04	3.79	3.37
	Seed 2	12.62	12.43	6.57	6.21	5.61
	Placenta 1	12.67	12.48	6.34	6.11	5.23
	Placenta2	15.11	14.99	8.08	7.77	6.70
	Flesh 1	8.91	8.78	5.23	4.84	4.26
	Flesh 2	8.35	8.29	5.32	5.15	4.65
	Total	135.13	131.19	78.60	74.82	63.81

751

752 **Table 1.** Summary of the read numbers (in million reads) along the filtering steps in the small

753 RNA libraries. Filter 1; removal of reads with low abundance (1-2). Filter 2; removal of

754 tRNA, rRNA, snRNA and snoRNA and low complexity sequences. Aligned reads are
755 matched to the pepper genome.
756

Figure legends

Figure 1. Developmental stages of the sweet pepper cv. “Fehér özön” investigated in this study. A. The two selected fruit developmental stages at 28 and 40 days after anthesis (DAA). Grid=1 cm. B. Vertical and horizontal sections of expanding ovary (fruit) of sweet pepper showing the composing tissue parts as indicated. The seeds, placenta and flesh were isolated and used for RNA extractions separately.

Figure 2. Size distribution of small RNA reads of two biological repeats of the different pepper fruit tissues as indicated. The diagram shows the small RNA size composition of each library. Biological repeats are displayed next to each other. Numbers at the bottom indicate the length if smRNAs.

Figure 3. Novel miRNA representatives**. A.** Read alignment diagram for the selected miRNA precursors. Mature strand highlighted with green and star strand with red. **B.** Predicted secondary structure of pre-miRNAs (Mfold). Mature miRNA and miRNA* sequence are highlighted.

Figure 4. Small RNA Northern blot analyses of selected conservative (**A**) and newly predicted (**B**) miRNAs. RNA samples, seed, placenta and flesh, representing developmental stages at 28 and 40 days after anthesis (DAA) were hybridized with probes detecting miRNAs as indicated. Upper panels are schematic representation of NGS read counts (read/million). Middle panels show the results of small RNA northern blots. Upper panels show the level of rRNAs serving as loading control. The same membrane was used for hybridization of mir171, mir396 and mir159,Can-mir-f13 specific probes.

Figure 5. Differential expression of conserved miRNAs visualized on MA plots between different tissue types at 28 DAA (**A**), at 40 DAA (**B**) and between the same tissue types (**C**). On the X axis we showed the mean of normalized reads (calculated by DESeq2). On Y axis there is the quantity changes (final value/initial value), where 1 means the final value is the 1× (100%) of the initial value, so the two expression is equal. We considered those changes as significant (red crosses), where the adjusted p-value was less than 0.05. Gray dots showing the not significant changes.

Figure 6. Differential expression of selected known miRNAs visualized on MA plots between different tissue types at 28 DAA (**A**), at 40 DAA (**B**) and between the same tissue types (**C**). On the X axis we showed the mean of normalized reads (calculated by DESeq2). On Y axis there is the quantity changes (final value/initial value), where 0.1 means the final value is the 0,1× (10%) of the initial value, so its expression was ten times lower. We considered those changes as significant (red crosses), where the adjusted p-value was less than 0.05. Gray dots showing the not significant changes.

Figure 7. Differential expression of novel miRNAs visualized on MA plots between different tissue types at 28 DAA (**A**), at 40 DAA (**B**) and between the same tissue types (**C**). On the X axis we showed the mean of normalized reads (calculated by DESeq2). On Y axis there is the quantity changes (final value/initial value), where 1 means the final value is the 10× (1000%) of the initial value, so it showed ten times higher expression. We considered those changes as significant (red crosses), where the adjusted p-value was less than 0.05. Gray dots showing the not significant changes.

Figure 8. Accumulation of AGO1 protein and miR168 in pepper fruit tissues. RNA and protein samples of seed, placenta and flesh, representing developmental stages at 28 and 40 days after anthesis (DAA), were used for small RNA northern blot analyses and Western

803 blotting. Upper panel is the schematic representation of miR168 NGS read counts
804 (read/million). Middle panels show the results miR168 northern blots. EtBr stained rRNAs
805 serve as loading control. Bottom panels show Western blot analyses of total protein extracts
806 for AGO1 accumulation. Equal loading was verified by Ponceau staining of the membrane
807 after western blotting.

808 **Figure 9.** Differential expression of non-miRNA overlapping sRNA clusters visualized on
809 MA-plots. Normalization and differential expression analysis was performed with DESeq2.
810 Clusters showing at least 2-fold abundance change with an adjusted p-value < 0.05 were
811 considered as significantly differentially expressed and colored red on the plots. Black dots
812 represent non-significantly changed clusters.

813 **Figure 10.** Expression pattern and sequence length composition of 24-nt-long hc-
814 siRNA/ohpRNA-producing clusters. Heat map showing the log2 transformed expression
815 levels (Read Per Million) of the 50 most abundant clusters (A). Sequence length composition
816 of the 50 most abundant clusters (B). Northern blot analyses of the expression of Cluster
817 297979 and Cluster 461037 (C).

818

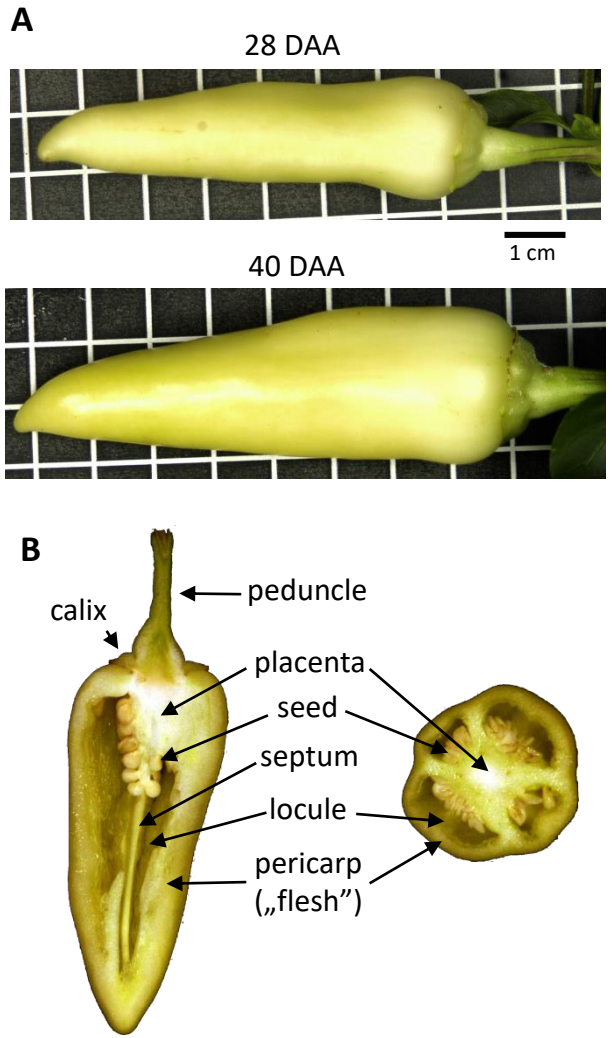

A
28 DAA
1 cm
40 DAA
B
calix
peduncle
placenta
seed
septum
locule
pericarp
(„flesh”)

Figure 1.

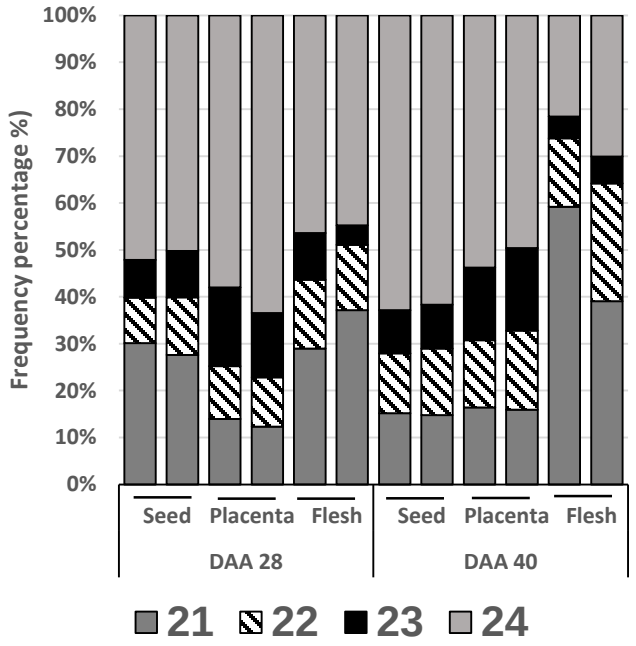

100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Frequency percentage %)
Seed
Placenta
Flesh
Seed
Placenta
Flesh
DAA 28
DAA 40
21
22
23
24

Figure 2.

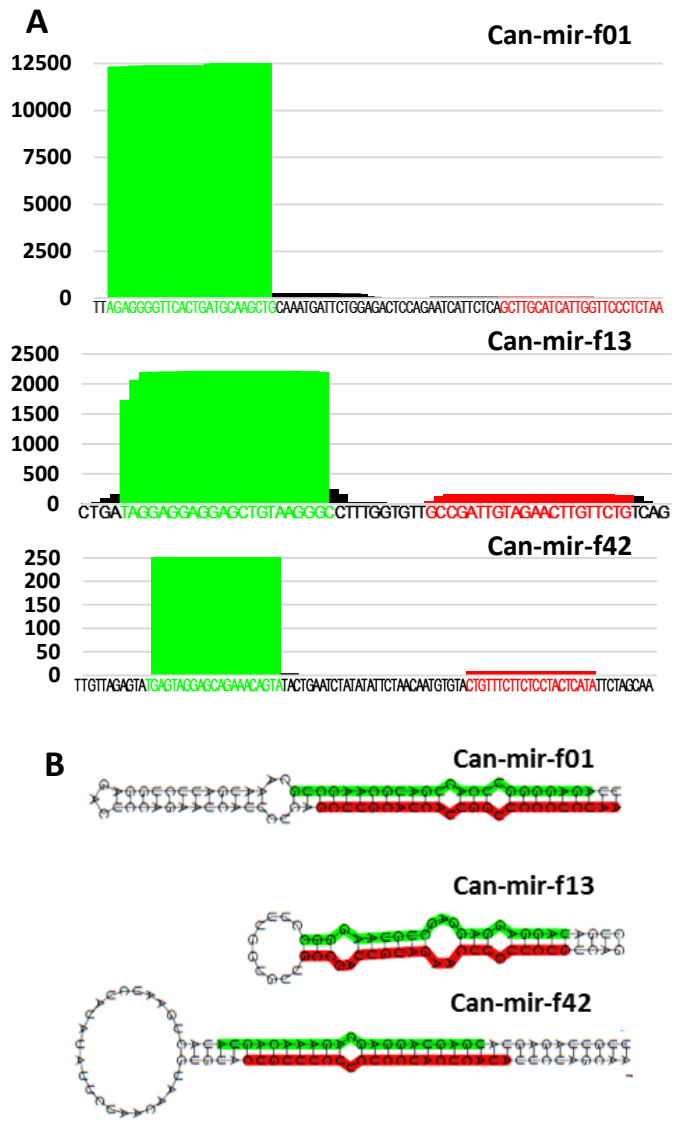

A
Can-mir-f01
12500
10000
7500
5000
2500
0
TTAGAGGGTTCACTGATGCAAGCTGCAAATGATTCTGGAGACTCCAGAATCATTCTCAGCTTGCATCATTGGTTCCCTCTAA
Can-mir-f13
2500
2000
1500
1000
500
0
CTGATAGGAGGAGGAGCTGTAAGGGCCTTTGGTGTTGCCGATTGTAGAACTTGTTCTGTCAG
Can-mir-f42
250
200
150
100
50
0
TTGTTAGAGTATGAGTAGGAGCAGAAACAGTATACTGAATCTATATATTCTAACAATGTGTACTGTTTCTTCTCCTACTCATATTCTAGCAA
B
Can-mir-f01
Can-mir-f13
Can-mir-f42

Figure 3.

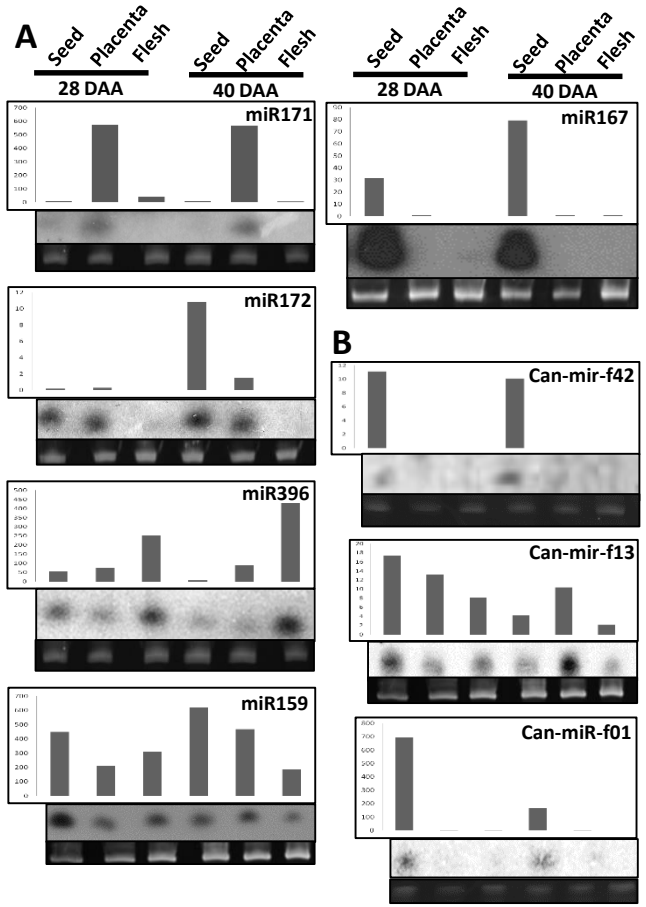

A
Seed
Placenta
Flesh
28 DAA
40 DAA
miR171
miR167
miR172
B
Can-mir-f42
miR396
Can-mir-f13
miR159
Can-miR-f01

Figure 4.

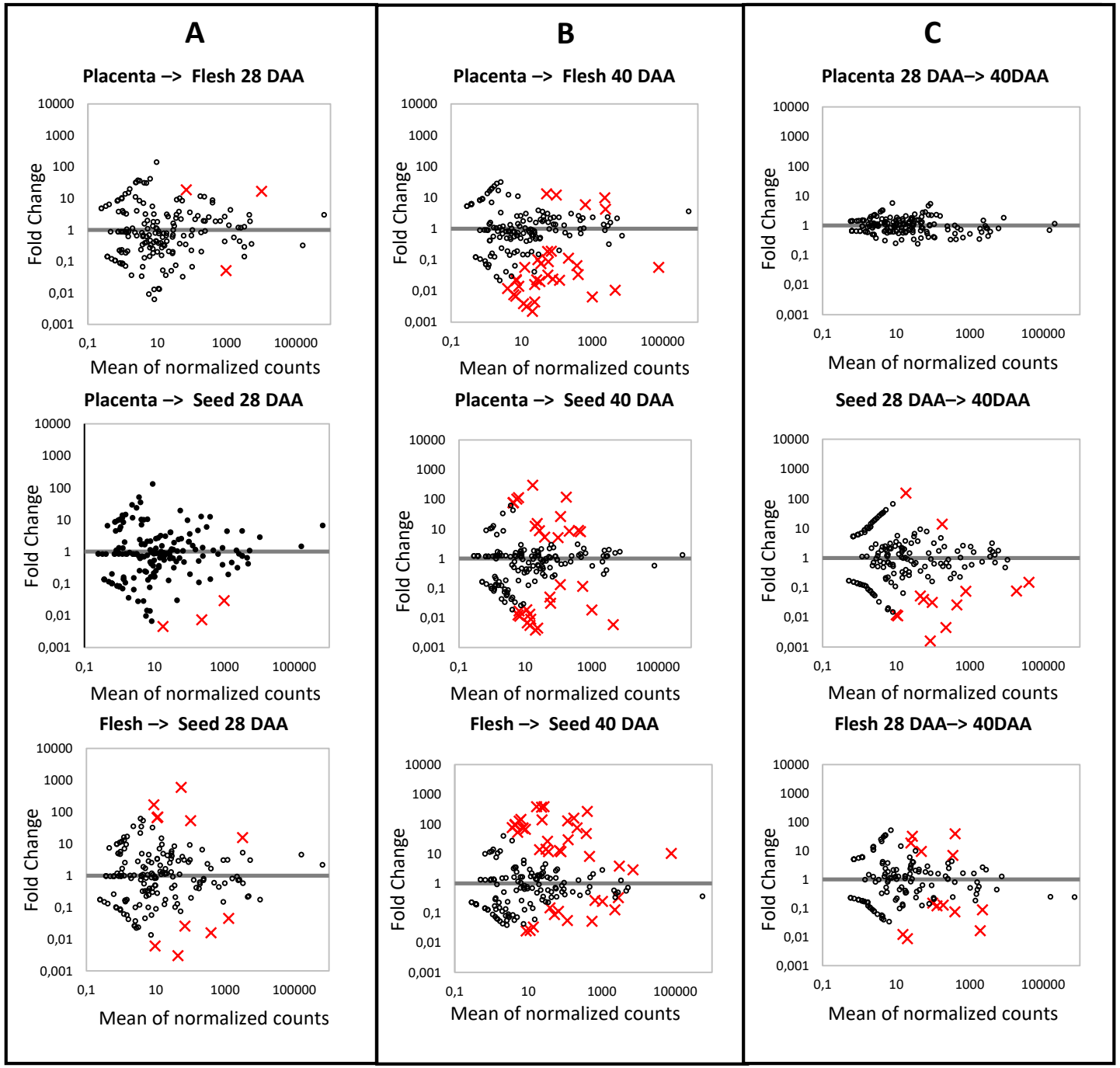

A
Placenta –> Flesh 28 DAA
Placenta –> Seed 28 DAA
Flesh –> Seed 28 DAA
B
Placenta –> Flesh 40 DAA
Placenta –> Seed 40 DAA
Flesh –> Seed 40 DAA
C
Placenta 28 DAA–> 40DAA
Seed 28 DAA–> 40DAA
Flesh 28 DAA–> 40DAA
Fold Change
Mean of normalized counts
10000
1000
100
10
1
0,1
0,01
0,001
0,1
10
1000
100000

Figure 5.

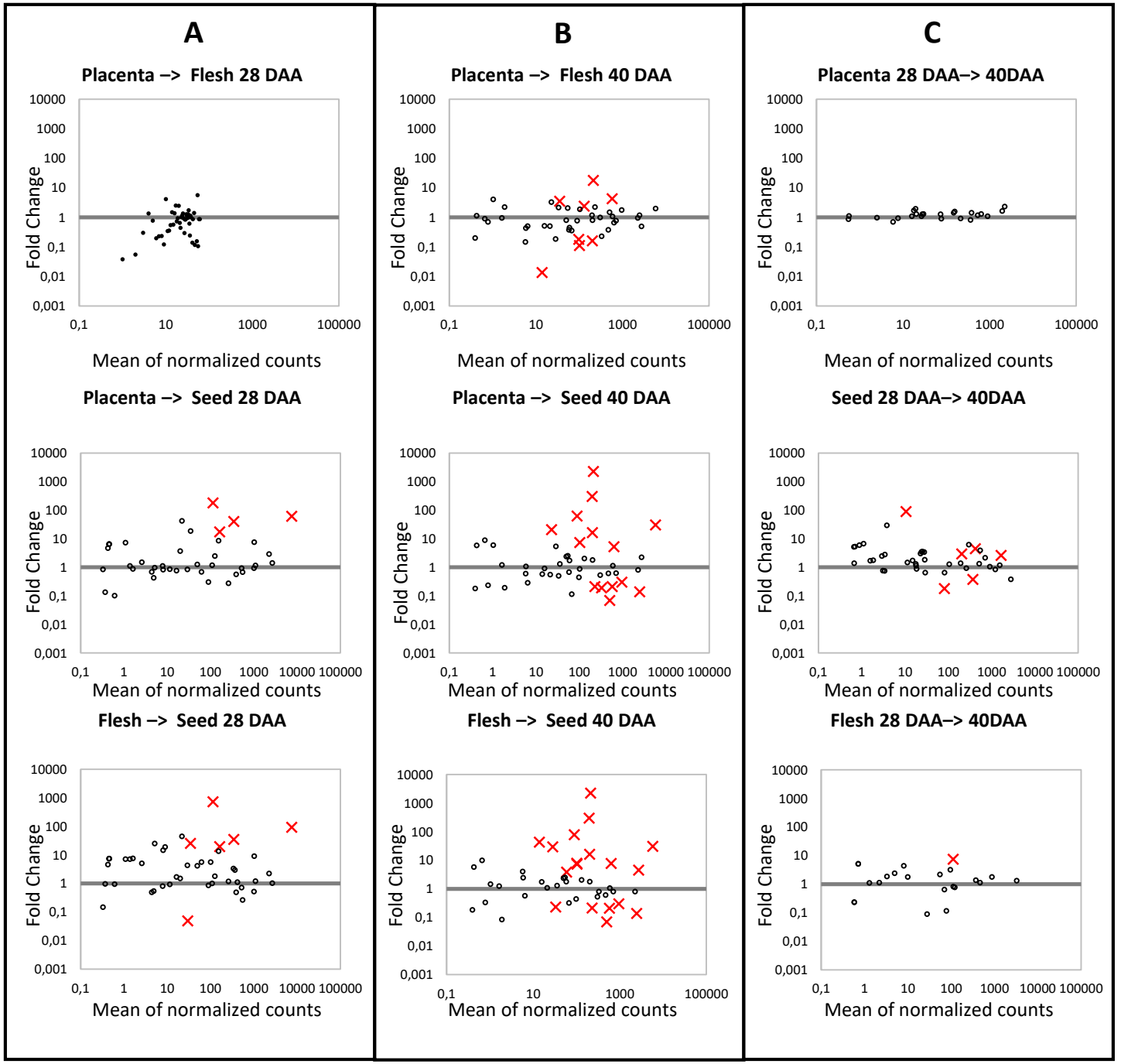

A
Placenta –> Flesh 28 DAA
Placenta –> Seed 28 DAA
Flesh –> Seed 28 DAA
B
Placenta –> Flesh 40 DAA
Placenta –> Seed 40 DAA
Flesh –> Seed 40 DAA
C
Placenta 28 DAA–> 40DAA
Seed 28 DAA–> 40DAA
Flesh 28 DAA–> 40DAA
Fold Change
Mean of normalized counts

Figure 6.

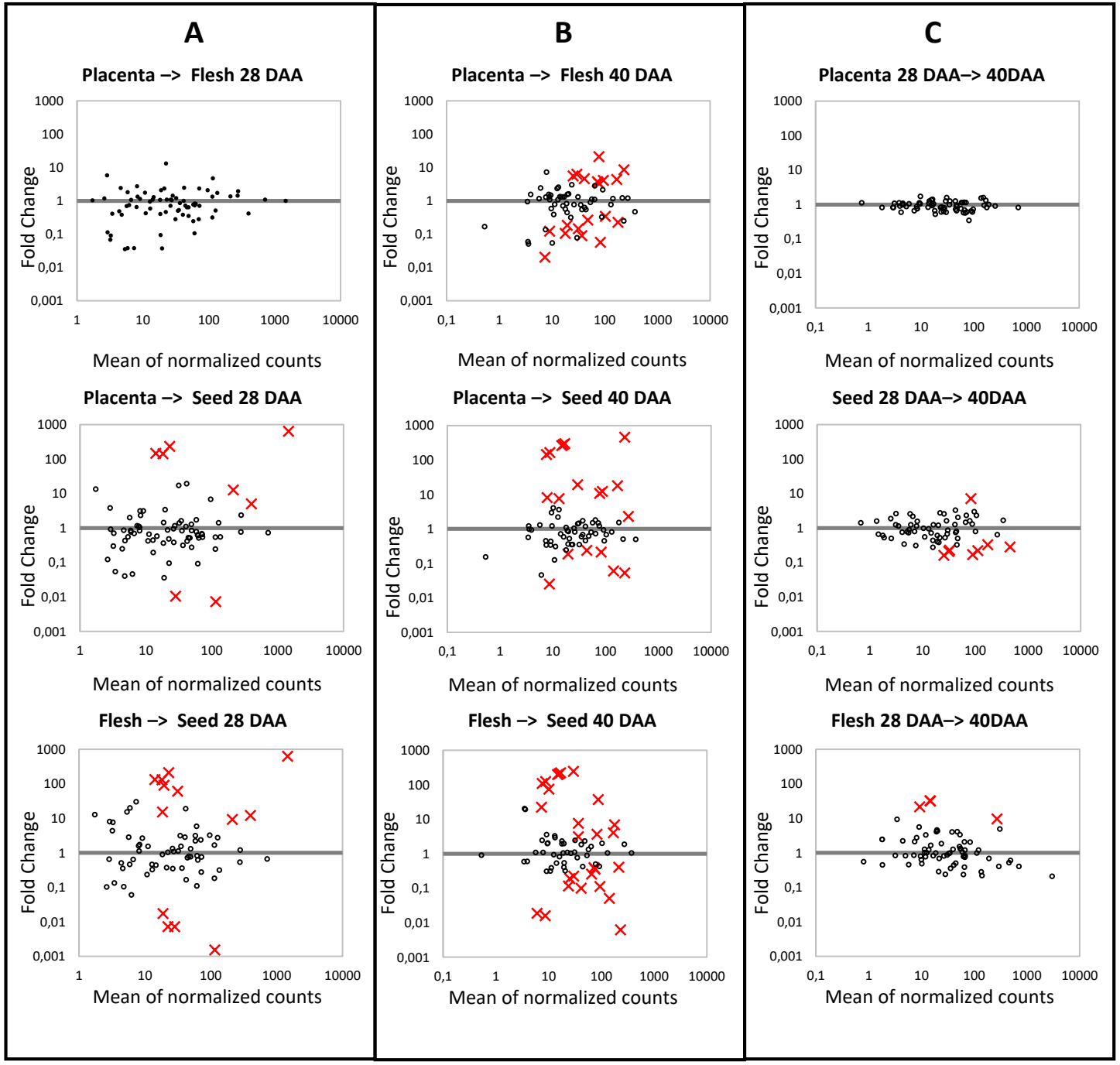

A
Placenta –> Flesh 28 DAA
Placenta –> Seed 28 DAA
Flesh –> Seed 28 DAA
B
Placenta –> Flesh 40 DAA
Placenta –> Seed 40 DAA
Flesh –> Seed 40 DAA
C
Placenta 28 DAA–> 40DAA
Seed 28 DAA–> 40DAA
Flesh 28 DAA–> 40DAA
Fold Change
Mean of normalized counts
1000
100
10
1
0,1
0,01
0,001
1
10
100
1000
10000
0,1

Figure 7.

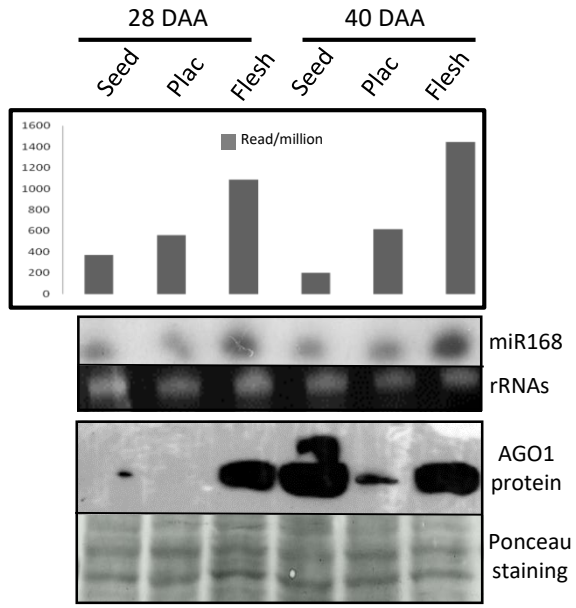

28 DAA
40 DAA
Seed
Plac
Flesh
Seed
Plac
Flesh
1600
1400
1200
1000
800
600
400
200
0
Read/million
miR168
rRNAs
AGO1
protein
Ponceau
staining

Figure 8.

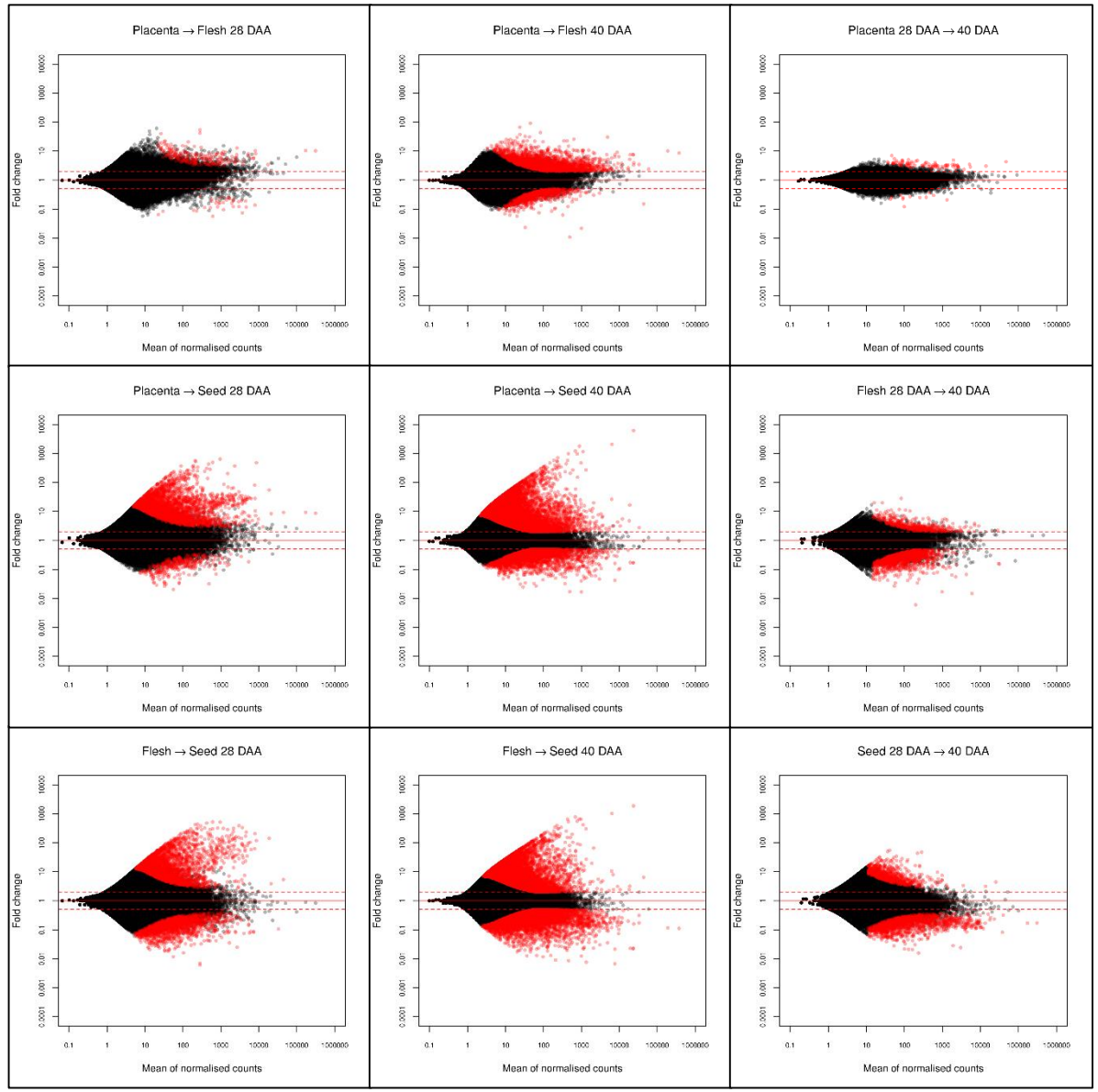

Placenta → Flesh 28 DAA
Placenta → Flesh 40 DAA
Placenta 28 DAA → 40 DAA
Placenta → Seed 28 DAA
Placenta → Seed 40 DAA
Flesh 28 DAA → 40 DAA
Flesh → Seed 28 DAA
Flesh → Seed 40 DAA
Seed 28 DAA → 40 DAA
Fold change
Mean of normalised counts

Figure 9.

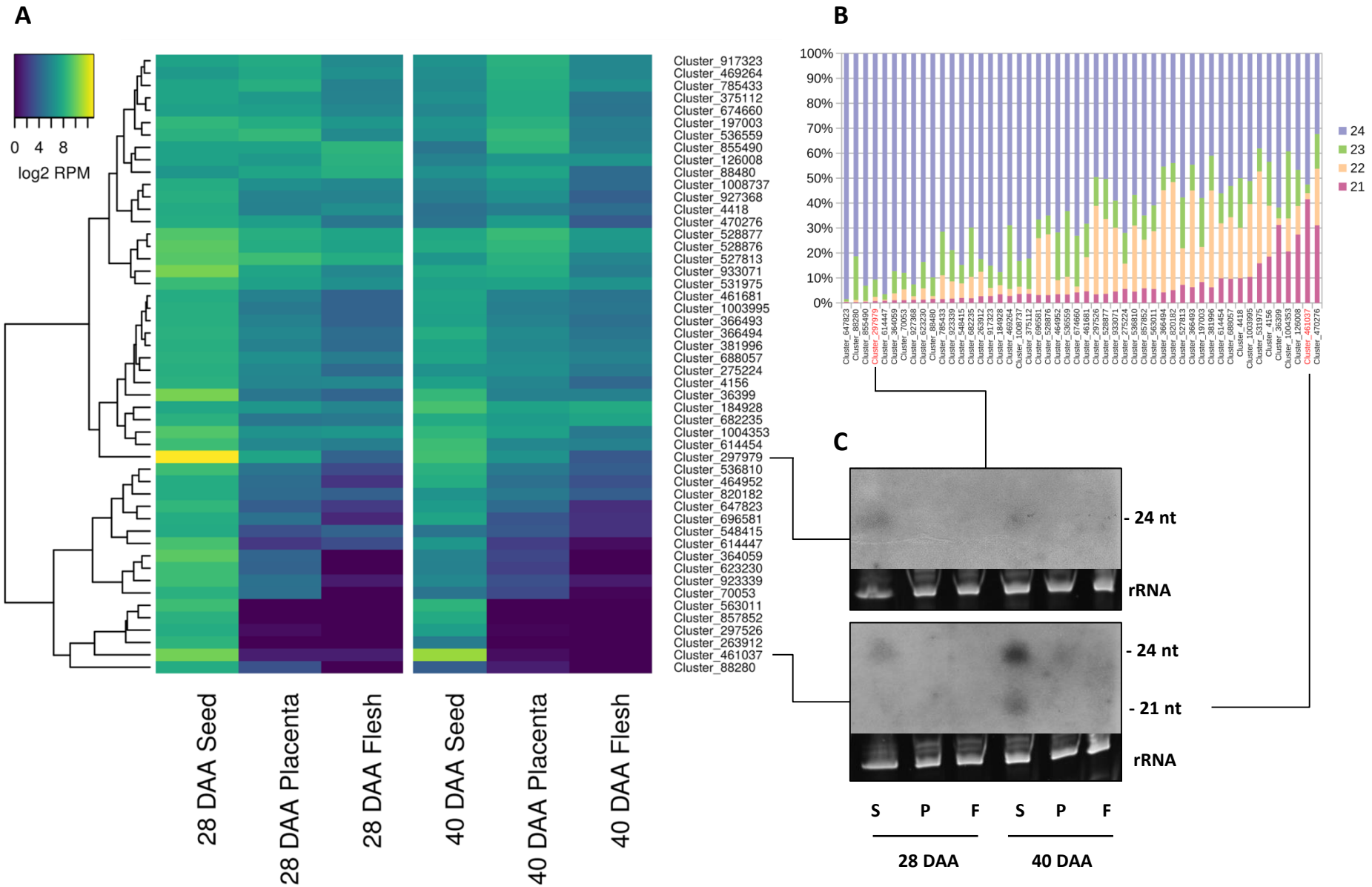

A
0 4 8
log2 RPM
Cluster_917323
Cluster_469264
Cluster_785433
Cluster_375112
Cluster_674660
Cluster_197003
Cluster_536559
Cluster_855490
Cluster_126008
Cluster_88480
Cluster_1008737
Cluster_927368
Cluster_4418
Cluster_470276
Cluster_528877
Cluster_528876
Cluster_527813
Cluster_933071
Cluster_531975
Cluster_461681
Cluster_1003995
Cluster_366493
Cluster_366494
Cluster_381996
Cluster_688057
Cluster_275224
Cluster_4156
Cluster_36399
Cluster_184928
Cluster_682235
Cluster_1004353
Cluster_614454
Cluster_297979
Cluster_536810
Cluster_464952
Cluster_820182
Cluster_647823
Cluster_696581
Cluster_548415
Cluster_614447
Cluster_364059
Cluster_623230
Cluster_923339
Cluster_70053
Cluster_563011
Cluster_857852
Cluster_297526
Cluster_263912
Cluster_461037
Cluster_88280
28 DAA Seed
28 DAA Placenta
28 DAA Flesh
40 DAA Seed
40 DAA Placenta
40 DAA Flesh
B
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
24
23
22
21
C
- 24 nt
rRNA
- 24 nt
- 21 nt
rRNA
S P F S P F
28 DAA
40 DAA

Figure 10.