University of Groningen

Computational Methods for High-Throughput Small RNA Analysis in Plants
Monteiro Morgado, Lionel

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2018

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Monteiro Morgado, L. (2018). Computational Methods for High-Throughput Small RNA Analysis in Plants. [Groningen]: University of Groningen.

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
CHAPTER 6

Small RNA reflects grandparental environments
in apomictic dandelion

Lionel Morgado, Veronica Preite, Carla Oplaat, Sarit Anava, Julie Ferreira de Carvalho, Oded Rechavi, Frank Johannes and Koen JF Verhoeven

Abstract

Plants can show long-term effects of environmental stresses and in some cases a stress “memory” has been reported to persist across generations, potentially mediated by epigenetic mechanisms. However, few documented cases exist of transgenerational effects that persist for multiple generations and it remains unclear if or how epigenetic mechanisms are involved. Here, we show that the composition of small regulatory RNAs in apomictic dandelion lineages reveals a footprint of drought stress and salicylic acid treatment experienced two generations ago. Overall proportions of 21 and 24 nt RNA pools were shifted due to grandparental treatments. While individual genes did not show strong up- or downregulation of associated sRNAs, the subset of genes that showed the strongest shifts in sRNA abundance was significantly enriched for several GO terms including stress-specific functions. This suggests that a stress-induced signal was transmitted across multiple unexposed generations leading to persistent changes in epigenetic gene regulation.

Molecular Biology and Evolution (2017)
6.1 Introduction

Stress exposure triggers responses that are mediated by changes in gene regulation [273–275]. In plants, some responses to environmental stresses are long-lived. For instance, upon mild pathogen infection, plants can enter a “primed” state which is expressed as a quicker or more vigorous defense response upon a second infection later in life [276]. Similar defense-related induced effects, and also responses to other environmental triggers, have been demonstrated to persist into offspring generations in some cases [277–280]. Although several different mechanisms can underlie inherited environmental effects in plants [281], epigenetic mechanisms are considered prime candidates because of their potential for environmental sensitivity [76] and transgenerational stability [251]. Especially DNA methylation can be transgenerationally stable in plants and this mechanism is often proposed to mediate environmental effects that persist for multiple generations [80, 282–284]. However, empirical support for this hypothesis remains scarce [285]. Accumulating evidence indicates that regulatory small RNA (sRNA) also has a role in plant

Figure 6.1. Apomictic Dandelion lines generated for the current study. Generations 1 to 3 are marked by G1, G2 and G3. Each plant cartoon accounts for 4 replicates, except for the single initial common ancestor individual (G0). Lines subjected to sRNA sequencing are indicated with a blue box. Treatment sets: control (C), drought (D); salicylic acid (S).
Figure 6.2. Length composition for the read libraries. All sRNAs (A and D), mapped to annotated TE (B and E) and mapped to gene-annotated transcripts (C and F) (mean ± SE). Bottom panels (D, E, and F) are enlargements of top panels showing P values from permutation tests performed for 21 and 24 nt sRNA size classes. P values larger than 0.1 are labeled “NS” (not significant). Treatment groups (C: control; D: drought; S: salicylic acid) refer to grandparental treatments.

Figure 6.3. Spatial accumulation of 21 and 24 nt sRNA reads in gene-mapping transcripts. Lines represent density distributions of sRNA mapping location along the transcript. Each gene-mapping transcript was scaled to a length of 1000 bp and sRNA mapping positions (pooled replicates) inside each transcript were transformed accordingly. The counts for each transcript were afterwards collapsed into a single transcript model by calculating the average number of sRNA hits for each transcript coordinates across all length-normalized transcripts. Color code for treatment groups: control (green), drought (blue), and SA (red).
transgenerational effects. Indeed, sRNA biogenesis mutants in *A. thaliana* show compromised transgenerational herbivore resistance [255], suggesting that sRNA is required to sustain induced defense responses across generations. Changes in sRNA composition have been demonstrated in a number of species in response to heat [80, 286, 287], drought [288, 289], salinity [17, 287, 288, 290], cold, and osmotic stress [287]. In some cases, these sRNA alternations have been shown to persist in the offspring of stressed plants [80]. The mechanisms that maintain changes of sRNA across generations remain largely unclear but may be the result of feedback loops involving (transiently) heritable DNA methylation changes [291]. Here, we used apomictic dandelion (*Taraxacum officinale*) to test the impact of environmental stress on sRNA composition in unexposed offspring two generations after stress treatment. Due to apomictic (clonal seed) reproduction, dandelion offspring are considered genetic copies allowing for multi-generation experiments without confounding effects of genetic differences between samples. Apomixis in triploid dandelion involves formation of unreduced egg cells that develop parthenogenetically into embryos [292]. It is possible that the absence of fertilization in apomicts promotes the transgenerational stability of novel epigenetic variants, as has been observed under vegetative propagation [293], because nonsexual reproduction may partly bypass the extensive reprogramming that occurs during male gametogenesis and early embryogenesis [294]. In apomictic dandelion, first-generation offspring of stress-exposed plants have previously been demonstrated to show modified phenotypes and DNA methylation patterns, suggesting potential for environment-induced transgenerational epigenetic inheritance [283, 295].

### 6.2 Materials and methods

The experimental design is visualized in Figure 6.1. We grew first-generation (G1) plants under either drought stress, salicylic acid exposure (SA; a plant hormone that is involved in several processes including defense signaling in response to pathogens [296]), or under control conditions. Second (G2) and third (G3) generation apomictic progenies were obtained by single-seed descent for four replicate lineages per experimental group and were grown under common control conditions. sRNA was sequenced at generation G3 in four individual plants per experimental group.
Apomictic dandelion plants from a single lineage (*T. officinale hemicyclus*) were used [297]. G1 plants in the experiment derived through apomictic reproduction from a single G0 individual. Seedlings were distributed over three treatment groups: control, drought and salicylic acid. Plants within a group were randomized and placed in rows to ensure non-touching between the treatment groups. After four weeks of growth in a climate chamber (14 h light / 10 h dark, 18 °C / 14 °C, 60%) drought treatment started: during ten brief drought episodes over a period of four weeks, water was withheld until at least 80% of all individuals showed wilted leaves. At five weeks after seedling growth, SA treatment was applied once by spreading 0.5 ml of 10 mM SA solution over the surface of three leaves. The control group received no treatment. At eight weeks, plants were moved to a cold room for a 5-week vernalization period (continuous 4°C, 16 h light / 8 h dark), after which plants were placed on greenhouse benches and seeds were collected six weeks later. Using single-seed descent the subsequent two generations, G2 and G3, were grown in a common control environment under the same conditions as described for G1.

Leaf tissue was sampled from five weeks old G3 plants. Sixteen leaf discs of 8 mm in diameter were punched from one young and fully-developed leaf. Leaf discs were snap frozen in liquid nitrogen and stored in -80°C until usage. From liquid nitrogen-ground leaf tissue total RNA was extracted using 1 ml of Trizol (Ambion, Life technologies, The Netherlands) according to the manufacturer protocol but with an additional precipitation step with isopropanol and 3M sodium acetate (pH 5.2) and subsequent washing steps with ethanol. The final pellet was dissolved in 50 μl DNase/RNase-free water. Quality was checked on agarose gel electrophoresis and concentration on a NanoDrop 1000 spectrophotometer. For RNA sequencing library preparation the New England Biolabs NEBNext multiplex small RNA kite7300 kit (Ornat, Rechovot, Israel) was used with an initial 1 μg of RNA according to the manufacturer protocol, using barcoded primers and using T4 ligase truncated for 3’ ligation. To select for sRNA with a length of 20-30nt, cuttings from a E-Gel EX 4% Agarose gel (Invitrogen, Life technologies, Israel) were taken between 140-150nt size fraction (accounting for two adapter sequences of 60nt each). Bands were cleaned using MiniElute Gel Extraction kit (QIAGen, Eldan, Israel). After clean-up and a final Bioanalyzer quality check (Agilent Technologies, Eldan, Israel) using Agilent High Sensitivity DNA Kit, the 12 samples were pooled into a single sequencing library which was sequenced on two Illumina Hiseq2500 lanes (50bp, single-end).
6.3 Results

As no reference assembly currently exists for dandelion, we first assessed differences in sRNA composition between treatment groups and control using the total sRNA libraries. Using permutation tests based on random reshuffling of sample labels when comparing G3 control samples to either G3 drought or SA samples, we found a significant reduction in the proportion of 24 nt sRNAs after grandparental SA treatment (bootstrap test, \(P=0.035\)), and also marginally significant shifts in 24 nt sRNAs after grandparental drought treatment (bootstrap test, \(P=0.094\)) and in 21 nt sRNAs after grandparental SA treatment (bootstrap test, \(P=0.086\)) (Figure 6.2A and D). The most pronounced changes occurred for sRNA of size 24 nt whose relative abundance in the total sRNA population was reduced in both of the stress conditions compared with the control. Relative loss of TE-associated 24 nt sRNA has been reported for a variety of biotic and abiotic stressors [76, 86, 298]. These changes appear to be mainly the result of hypomethylation and loss of RdDM targeting of transposable element (TE) sequences [229, 299]. To understand the stress-induced shifts in 21 and 24 nt sRNA composition in more detail, we took advantage of a recent TE database that was generated based on de novo clustering of repetitive sequences from the T. officinale genome [300]. We aligned sRNAs to these TE sequences, and compared the relative size abundance across conditions. A loss of 24 nt sRNAs was also observed in these TE-annotated sequences, at least after drought stress (\(P=0.052\), Figure 6.2B and E). Loss of 24 nt TE-associated sRNAs is typically accompanied by gains in 21 nt sRNAs due to an increase in the transcription of precursors for this class of sRNA [76, 86]. Consistent with this, the loss of 24 nt TE-mapping sRNAs after drought stress co-occurred with an increase in 21 nt sRNAs, although this increase was not statistically significant (\(P=0.139\), Figure 6.2E). It is unclear if and how such sRNA shifts impact gene expression. Previous studies have reported that TEs proximal to or overlapping genes can affect transcription under stress, probably as a result of RdDM-mediated DNA methylation loss [76, 167, 301–303]. Another mechanism by which sRNA can affect gene expression is through trans-acting posttranscriptional modifications [17]. We recently assembled the complete transcriptome of dandelion [304]. Over 13,500 genes could be annotated by homology with A. thaliana. We aligned our sRNA libraries to these transcriptomes. On average about 53% of the reads from each library met our quality control criteria and could be successfully aligned. Similar to TE-
Per-transcript analysis of differential sRNA abundance. DESeq2 results for: all consensus sRNA (sRNA shared by all replicates from at least one of the treatments), TE-mapping sRNA and gene-mapping sRNA. A FDR threshold of 0.1 was applied after multiple testing corrections.

| Group          | Comparison | sRNA length (nt) | #Instances tested | #Up-regulated | #Down-regulated |
|----------------|------------|------------------|-------------------|---------------|-----------------|
| Consensus set  | All        | 21               | 569268            | 3             | 0               |
|                | 24         | 102759           | 235625            | 2             | 0               |
|                | D/C        | All              | 545459            | 15            | 12              |
|                |            | 21               | 99072             | 5             | 8               |
|                | S/C        | 24               | 227383            | 0             | 6               |
| TEs            | All        | 98               | 35880             | 0             | 5               |
|                | D/C        | 21               | 27153             | 0             | 0               |
|                | S/C        | 24               | 27730             | 0             | 1               |
| Gene-mapping   | All        | 98               | 36202             | 0             | 0               |
| transcripts    | D/C        | 21               | 27975             | 0             | 0               |
|                | S/C        | 24               | 28683             | 0             | 0               |

Figure 6.4. Overlap between GO terms significantly enriched in the list of most downregulated and upregulated genes due to grandparental stress treatment. The downregulated set comprises the 5% transcripts with strongest reduction in mapped sRNA, when comparing stress-derived against control. The upregulated set is composed by the 5% transcripts with the strongest increase in mapped sRNA, comparing stress to control sRNA. Numbers for significantly enriched GO terms are shown as Venn diagrams for effects of grandparental drought and grandparental SA treatment for 21 nt and 24 nt sRNA and for all length classes between 18 and 30 nt.
aligned sRNAs, the relative abundance of transcriptome-aligned 24 nt sRNAs was reduced after grandparental stress treatments (most clearly after grandparental SA treatment: bootstrap test, \(P=0.04\); see Figure 6.2C and F), but now also a reduction in the relative abundance of 21 nt sRNA was indicated (bootstrap test, \(P=0.068\); Figure 6.2F). The observed loss of transcriptome-aligned 24 nt sRNA was enigmatic, as 24 nt sRNA are typically depleted in genic sequences. To explore this issue in more detail, we studied the density distribution of 24 nt sRNA along our annotated transcripts. Our analysis shows that 24 nt sRNA mostly mapped towards the 5’ and 3’ flanks of the genes, suggesting vestiges of a sRNA signal that originates from sequences outside of gene bodies, such as promoters or intergenic regions (Figure 6.3). In contrast, 21 nt sRNA showed a peak density more toward the center of gene bodies. The distributional patterns reported here resemble previously reported genic signatures of sRNA abundance in well-annotated genomes such as maize [298, 305], A. thaliana [76], and rice [306]. The relative decrease in transcript-associated 24 nt sRNA after stress exposure may suggest a loss of methylation in gene flanking regions (and possibly also in TE sequences within genes) and consequent gene expression upregulation. However, a quality genome reference assembly will be required to test this hypothesis.

Together, the specific changes in sRNA profiles that we observed are in line with previous observations in stress-exposed plants, but our results indicate that the stress-associated sRNA footprint is maintained transgenerationally for at least two unexposed generations after the stress treatment. In order to identify specific genes that show different sRNA abundance comparing control and stress treatments, we performed differential analysis using DESeq2. DESeq2 uses negative binomial generalized linear models to statistically test each gene for a difference between experimental groups in the number of (sRNA) reads mapping to that gene [234]. We applied DESeq2 to different sRNA lengths: 21 nt, 24 nt sRNAs and for all length classes combined (18–30 nt). After adjusting for multiple testing (FDR=0.1), our results showed virtually no significant sRNA enrichment or depletion at individual genes (Table 6.1). This is consistent with an observed overall high similarity between individual sRNA libraries, both within and between experimental groups. We argue that the induced and transgenerationally inherited sRNA effects are subtle and may not be readily detectable using our approach that involved sRNA sequencing of individual plants, not pooled samples. We therefore focused, instead, on sets of genes that were either most depleted or enriched for sRNA in the stress groups and we tested if these gene sets were
overrepresented for specific GO-terms. Using Fisher’s Exact tests our analysis revealed a significant overrepresentation for hundreds of GO categories, depending on grandparental treatment and sRNA length class (our significance testing included evaluation against a bootstrapping-derived null distribution of enrichment to account for potential baseline biases in the dandelion transcriptome when compared with the Arabidopsis reference gene set. Subsampling was repeated 1000 times picking a number of random transcripts equal to the top 5% of the respective treatment-control set. We considered the enrichment to be significant if the FDR-adjusted p-values from the topGO enrichment analyses exceeded the 95% confidence interval of the p-values from the bootstrapped subsamples). A large fraction of these significant GO-categories overlapped between the two stress treatments, suggesting a generalized stress response. The 5% of genes that were most depleted or enriched for 21 nt sRNA were significantly enriched for about 400–500 GO categories in both the control-drought comparison and in the control-SA comparison. The 5% of genes that were most depleted for 24 nt sRNA were enriched for many more GO terms than the 5% of genes that showed the strongest increase in associated 24 nt sRNA, suggesting a strong biological signal in the relaxation of 24 nt-based gene silencing after grandparental stress.

We searched the list of significantly enriched GO terms for specific keywords that are associated with the grandparental stresses: “water” and “drought” for drought treatment, “salicylic” and “hormone” for SA treatment and “response to stress”, “abiotic stimulus” and “wounding” for stress treatments in general (see Figure 6.5). For instance, the GO term 0006950 (“response to stress”) showed strong statistical evidence for enrichment both after drought and after SA treatment, pointing to an active stress memory. For all key words, except “drought” for which no significantly enriched GO terms were detected, significantly enriched GO terms were found in both stress treatments, suggesting that these GO terms reflect a general stress response rather than a treatment-specific response. However, two SA-related GO categories (GO term 0009862: systemic acquired resistance, salicylic-mediated signaling pathway; and GO term 0009914: hormone transport) were affected only in the SA set, indicating a more treatment-specific pattern.
6.4 Discussion

In summary, it is well known that stress responses can be mediated by changes in sRNA-associated gene silencing. Our results suggest that this regulation may persist for several generations after stress. sRNA-based multi-generational inheritance of environmental stress has been previously demonstrated in some animal systems (e.g. [307, 308]) where underlying mechanisms of sRNA inheritance are at least partly different from plants. Although effects on gene expression remain to be evaluated, our study is to our knowledge the first demonstration in plants of modified sRNA two generations removed from the stress trigger. Our results show no clear statistically significant effects on individual genes, which may be due to low sequencing depth of the libraries, or lack of sensitivity of our differential analysis. However, we were able to uncover a sRNA signal among genes involved in stress-related functions. This illustrates that an epigenetic signal travelled between generations preserving footprints of grandparental stress, and that this memory implicates genes that are known to be involved in stress responses. Although we did not explore the nature of the transgenerationally inherited epigenetic signal, this signal could be a stress-induced change in TE-associated DNA methylation, which in plants can be stably inherited and can trigger RNA-mediated gene expression changes in offspring [291].
Figure 6.5. Distribution for sRNA fold change in gene-mapping transcripts after grandparental drought stress (A) and salicylic acid (B) against control. Bar plots show p values of GO term enrichment tests (blue: enrichment test in set of genes with reduced sRNAs after stress; orange: enrichment test in set of genes with increased sRNAs after stress), in the case of the drought (C) and the salicylic acid (D) sets. Grey bars indicate P values obtained from random bootstrapping (absence of enrichment), which can be affected by biases in the dandelion reference transcriptome in comparison to the Arabidopsis reference gene set. Error bars indicate 95% bootstrap confidence intervals.