Small RNAs reflect grandparental environments in apomictic dandelion

Lionel Morgado¹*, Veronica Preite²*, Carla Oplaat², Sarit Anava³, Julie Ferreira de Carvalho², Oded Rechavi³, Frank Johannes⁴,⁵ and Koen J.F. Verhoeven²

* Equal contributions

1 Groningen Bioinformatics Centre, University of Groningen, Nijenborgh 7, 9747 AG Groningen, The Netherlands

2 Department of Terrestrial Ecology, Netherlands Institute of Ecology (NIOO-KNAW), Droevendaalsesteeg 10, 6708 PB Wageningen, The Netherlands

3 Department of Neurobiology, Wise Faculty of Life Sciences and Sagol School of Neuroscience, Tel Aviv University, Tel Aviv 69978, Israel

4 Population epigenetics and epigenomics, Technical University of Munich, Liesel-Beckmann-Strasse 2, Freising 85354, Germany.

5 Institute for Advanced Study, Technical University of Munich, Lichtenbergstrasse 2a, Garching 85748, Germany

Corresponding author: Koen Verhoeven (k.verhoeven@nioo.knaw.nl)
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 21nt and 24nt 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.
Stress exposure triggers responses that are mediated by changes in gene regulation (Heil 2002; Shao et al. 2008; Cramer et al. 2011). 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 (Conrath et al. 2002). Similar defense-related induced effects, and also responses to other environmental triggers, have been demonstrated to persist into offspring generations in some cases (Agrawal 2002; Mandal et al. 2012; Slaughter et al. 2012; Wang et al. 2016).

Although several different mechanisms can underlie inherited environmental effects in plants (Crisp et al. 2016), epigenetic mechanisms are considered prime candidates because of their potential for environmental sensitivity (Dowen et al. 2012) and transgenerational stability (Cortijo et al. 2014). Especially DNA methylation can be transgenerationally stable in plants and this mechanism is often proposed to mediate environmental effects that persist for multiple generations (Boyko et al. 2007; Boyko et al. 2010; Verhoeven et al. 2010; Ou et al. 2012; Bilichak et al. 2015). However, empirical support for this hypothesis remains scarce (Pecinka and Scheid 2012).

Accumulating evidence indicates that regulatory small RNAs (sRNAs) also have a role in plant transgenerational effects. Indeed, sRNA biogenesis mutants in *A. thaliana* show compromised transgenerational herbivore resistance (Rasmann et al. 2012), suggesting that sRNAs are required to sustain induced defense responses across generations. Changes in sRNA composition have been demonstrated in a number of species in response to heat (Ito et al. 2011; Bilichak et al. 2015; Song et al. 2016), drought (Matsui et al. 2008; Tricker et al. 2012), salinity (Borsani et al. 2005; Matsui et al. 2008; Ding et al. 2009; Song et al. 2016), cold and osmotic stress (Song et al. 2016). In some cases, these sRNA alternations have been shown to persist in the offspring of stressed plants (Bilichak et al. 2015). The mechanisms that maintain changes
of sRNAs across generations remain largely unclear but may be the result of feed-back loops involving (transiently) heritable DNA methylation changes (Wibowo et al. 2016).

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 (Bicknell and Koltunow 2004). 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 (Ong-Abdullah et al. 2015), because non-sexual reproduction may partly bypass the extensive reprogramming that occurs during male gametogenesis and early embryogenesis (Kawashima and Berger 2014). 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 (Verhoeven et al. 2010; Verhoeven and van Gurp 2012).

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; Vicente and Plasencia 2011), 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 (Supplementary text: S1). sRNAs were sequenced at generation G3 in four individual plants per experimental group (Supplementary text: S1-S2). 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 (Supplementary text: S3), 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) (Fig. 1A and 1D). The most pronounced changes occurred for sRNAs of size 24 nt whose relative abundance in the total sRNA population was reduced in both of the stress conditions compared to the control. Relative loss of TE-associated 24nt sRNA has been reported for a variety of biotic and abiotic stressors (Dowen et al. 2012; Lunardon et al. 2016; McCue et al. 2012). These changes appear to be mainly the result of hypomethylation and loss of RdDM targeting of transposable element (TE) sequences (Tran et al. 2005; McCue et al. 2013).

To understand the stress-induced shifts in 21 nt 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 (Ferreira de Carvalho et al. 2016a) (Supplementary text: S2). 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, Fig. 1B and 1E). 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 (Dowen et al. 2012; McCue et al. 2012). 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, Fig. 1E).

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 (Lister et al. 2008; Hollister and
Another mechanism by which sRNAs can affect gene expression is through trans-acting post-transcriptional modifications (Borsani et al. 2005). We recently assembled the complete transcriptome of dandelion (Ferreira de Carvalho et al. 2016b). Over 13500 genes could be annotated by homology with *A. thaliana*. We aligned our sRNA libraries to these transcriptomes (Supplementary text: S2). On average about 53% of the reads from each library met our quality control criteria and could be successfully aligned. Similar to TE-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 Fig. 1C and 1F), but now also a reduction in the relative abundance of 21 nt sRNAs was indicated (bootstrap test, p=0.068; Fig. 1F).

The observed loss of transcriptome-aligned 24 nt sRNAs was enigmatic, as 24nt sRNA are typically depleted in genic sequences. To explore this issue in more detail, we studied the density distribution of 24nt sRNA along our annotated transcripts. Our analysis shows that 24 nt sRNAs 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 (Fig. 2). 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 (Gent et al. 2013; Lunardon et al. 2016), *A. thaliana* (Dowen et al. 2012) and rice (Li et al. 2012). 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...
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 (Love et al. 2014). We applied DESeq2 to different sRNA lengths: 21nt, 24nt sRNAs and for all length classes combined (18 to 30nt). After adjusting for multiple testing (FDR=0.10), our results showed virtually no significant sRNA enrichment or depletion at individual genes (Supplementary table: S4). This is consistent with an observed overall high similarity between individual sRNA libraries, both within and between experimental groups (Supplementary text: S2). 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 (Supplementary text: S5; 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 to the Arabidopsis reference gene set). 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 sRNAs 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 sRNAs were enriched for many more GO terms than the 5% of genes that
showed the strongest increase in associated 24 nt sRNAs, 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 Fig. 3). 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.

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. Gapp et al. 2014; Rechavi et al. 2014) 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 sRNAs 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 (Wibowo et al. 2016).

Acknowledgments and funding information

Keygene N.V., Wageningen, kindly provided *T. officinale* BAC sequences that we used for
preliminary analysis of TE-mapping sRNAs. Seeds from the *T. officinale hemicyclum* lineage
were kindly provided by Jan Kirschner from the CAS Institute of Botany, Pruhonice. This work
was supported by the Netherlands Organisation for Scientific Research (grant numbers
864.10.008 and 884.10.003); and an ERC starting grant to O.R. (grant number 335624). F.J.
acknowledges support from the Technical University of Munich-Institute for Advanced Study
funded by the German Excellence Initiative and the European Union Seventh Framework
Programme under grant agreement #291763. sRNA data generated for this study are deposited
in the NCBI Sequence Read Archive (SRA study SRP096310, BioProject accession number
PRJNA360587).

Literature Cited

Agrawal AA. 2002. Herbivory and maternal effects: Mechanisms and consequences of

Bicknell RA, Koltunow AM. 2004. Understanding apomixis: Recent advances and remaining


McCue AD, Nuthikattu S, Reeder SH, Slotkin RK. 2012. Gene Expression and Stress Response Mediated by the Epigenetic Regulation of a Transposable Element Small RNA. Plos Genetics 8(2).


Figure legends

Figure 1. Length composition for the read libraries: all sRNAs (A and D), mapped to annotated TEs (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 21nt and 24nt 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 2. 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 averaged number of sRNA hits for each transcript coordinate across all length-normalized transcripts. Color code for treatment groups: control (green), drought (blue) and SA (red).

Figure 3. 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.
Figure 2

![Graphs showing sRNA accumulation for 21nt and 24nt sequences. Each graph represents the relative position in the gene from 5' to 3', with maximal (Max) and minimal (Min) values indicated.]
Figure 3

Test for GO enrichment

- Lower 5%
- Upper 5%
- Bootstrap (+/- 95% CI)

Selected GO categories:
- GO:0009414 | response to water deprivation
- GO:0009751 | response to salicylic acid
- GO:0009725 | response to hormone
- GO:0009755 | hormone-mediated signaling pathway
- GO:0009415 | response to water deprivation
- GO:0009752 | response to salicylic acid
- GO:0009726 | response to hormone
- GO:0009756 | hormone-mediated signaling pathway

GO:0009914 | hormone transport
GO:0006950 | response to stress
GO:0080134 | regulation of response to stress
GO:0080136 | priming of cellular response to stress
GO:009611 | response to wounding
GO:009628 | response to wounding