

RESEARCH ARTICLE

Open Access

Tissue-specific transposon-associated small RNAs in the gymnosperm tree, Norway spruce



Miyuki Nakamura^{*} , Claudia Köhler and Lars Hennig

Abstract

Background: Small RNAs (sRNAs) are regulatory molecules impacting on gene expression and transposon activity. MicroRNAs (miRNAs) are responsible for tissue-specific and environmentally-induced gene repression. Short interfering RNAs (siRNA) are constitutively involved in transposon silencing across different type of tissues. The male gametophyte in angiosperms has a unique set of sRNAs compared to vegetative tissues, including phased siRNAs from intergenic or genic regions, or epigenetically activated siRNAs. This is contrasted by a lack of knowledge about the sRNA profile of the male gametophyte of gymnosperms.

Results: Here, we isolated mature pollen from male cones of Norway spruce and investigated its sRNA profiles. While 21-nt sRNAs is the major size class of sRNAs in needles, in pollen 21-nt and 24-nt sRNAs are the most abundant size classes. Although the 24-nt sRNAs were exclusively derived from TEs in pollen, both 21-nt and 24-nt sRNAs were associated with TEs. We also investigated sRNAs from somatic embryonic callus, which has been reported to contain 24-nt sRNAs. Our data show that the 24-nt sRNA profiles are tissue-specific and differ between pollen and cell culture.

Conclusion: Our data reveal that gymnosperm pollen, like angiosperm pollen, has a unique sRNA profile, differing from vegetative leaf tissue. Thus, our results reveal that angiosperm and gymnosperm pollen produce new size classes not present in vegetative tissues; while in angiosperm pollen 21-nt sRNAs are generated, in the gymnosperm Norway spruce 24-nt sRNAs are generated. The tissue-specific production of distinct TE-derived sRNAs in angiosperms and gymnosperms provides insights into the diversification process of sRNAs in TE silencing pathways between the two groups of seed plants.

Keywords: Gymnosperm, Male gametophyte, Norway spruce, Small RNA, Transposable elements

Background

There are several different types of functional small RNAs (sRNAs) in animals and plants that differ in their biogenesis pathways, size classes, and functions. MicroRNAs (miRNAs), which are 21-nucleotides (nt) sRNAs derived from hairpin precursors, are the most common sRNA species generated in both plants and animals. Several miRNA sequences are conserved across land plants (reviewed in [1–3]), revealing their ancient functions and conservations. In contrast to miRNAs that mainly regulate protein-coding genes, small interfering RNAs (siRNAs) silence

transposable elements (TEs) and repeat sequences by either inducing degradation of TE transcripts via the post-transcriptional gene silencing (PTGS) pathway or inducing heterochromatin formation via the RNA-dependent DNA methylation (RdDM) pathway. sRNAs are prevalent to induce heterochromatin formation of their target sequences across kingdoms. In fission yeast *Schizosaccharomyces pombe* the RITS (RNA-induced transcriptional silencing) pathway confers siRNA-mediated heterochromatinization [4]. In animals, 24–30-nt Piwi-interacting RNAs (piRNAs) cause silencing of TEs by cleaving TE-derived transcripts in germline cells [5]. Likewise, in angiosperms, 24-nt small interfering RNAs (siRNAs) contribute to TE silencing through DNA methylation (reviewed in [6]).

* Correspondence: miyuki.nakamura@slu.se

Department of Plant Biology, Uppsala BioCenter, Swedish University of Agricultural Sciences and Linnean Center for Plant Biology, Uppsala, Sweden



The plant-specific RNA polymerase IV (Pol IV) and RNA polymerase V (Pol V) are involved in the RdDM pathway through 24-nt siRNAs, which are also called heterochromatic siRNA or repeat-associated siRNA [6, 7]. In flowering plants, 24-nt siRNAs occupy the predominant fraction of the sRNA population [8–13]. This 24-nt sRNA fraction is specifically composed of siRNAs that are derived from TEs [14, 15]. This strong positive association between 24-nt siRNAs and TEs allows us to use 24-nt siRNAs for TE annotations in both monocots and dicots [16].

Pol IV emerged after the divergence of algae and land plants [17]. However, in other land plants including gymnosperms, 24-nt sRNAs are not the predominant size class in vegetative tissue [11, 18, 19], although sRNAs are proposed to be involved in DNA methylation in moss and gymnosperms [20, 21]. Nevertheless, the factors involved in the RdDM pathway, such as RNA-DEPENDENT RNA POLYMERASE 2 (RDR2), DICER-LIKE 3 (DCL3), ARGONAUTE 4 (AGO4) and most of Pol IV and Pol V components, are conserved between angiosperms and gymnosperms [17, 21]. Indeed, gymnosperms have identifiable 24-nt sRNA populations in male cones and embryonic tissues [18, 21–25], revealing tissue-specific differences in sRNA production and possibly function in gymnosperms. More than half of the genomes in most gymnosperm species are occupied by TEs [18, 26]. Considering their very large genomes (4–30 Giga base pairs), the number of TEs is massive in spite of the limited amount of 24-nt sRNAs. This raises the question of which type of sRNA is involved in TE silencing in gymnosperms. It has been reported that the sRNA populations in angiosperm pollen are distinct from other tissues. *Arabidopsis* pollen accumulates TE-derived epigenetically activated siRNAs (easiRNAs) of 21–22-nt length [27]. In maize and rice anthers, 21-nt and 24-nt phased sRNAs (phasiRNAs) derived from intergenic regions make up a large fraction of the sRNA populations [28, 29].

In gymnosperms, it was reported that male cones have 24-nt sRNA [18, 24, 30]. However, it remains unclear whether those 24-nt sRNAs are derived from pollen. To investigate a possible regulatory association between sRNAs and TEs in different tissue types of gymnosperms, we generated and analyzed the sRNA profile of pollen. Moreover, we also investigate the sRNA from somatic embryonic callus, which is known to contain 24-nt sRNA fractions [21], along with needle samples representing a vegetative tissue of Norway spruce (*Picea abies*).

Results

Norway spruce pollen generates 24-nt sRNAs that are exclusively derived from TE sequences

To test whether gymnosperm pollen generates a specific population of TE-derived sRNAs, we harvested pollen from mature male cones of Norway spruce. We generated and sequenced sRNA libraries of pollen and

needles as vegetative tissue control (Additional file 1: Table S1). The sRNA size distribution strongly differed among the tissues (Fig. 1a, b); while needles had a large peak at 21-nt, pollen samples had two peaks at 21-nt and 24-nt length. When only considering non-redundant sequences, the 24-nt peaks in pollen were even more evident (Fig. 1b), consistent with previous observations in male cones [18, 24, 30]. We conclude that Norway spruce pollen generates tissue-specific 24-nt sRNAs.

To address the origin of sequences producing sRNAs, we assigned mapped reads to genomic features; genes and TEs. Gene regions generated only 21-nt sRNAs in both needles and pollen (Fig. 1c, e). For TE-derived sRNAs, needles exhibited a peak only at 21-nt, while pollen exhibited two peaks at both 21-nt and 24-nt (Fig. 1d, f). Thus, like in angiosperm pollen, TE sequences in Norway spruce pollen can produce both 24-nt and 21-nt sRNAs (Fig. 1f). Interestingly though, the sRNA size distributions in Norway spruce pollen are opposite of those in *Arabidopsis*, with a substantially higher fraction of 21-nt sRNAs accumulating in Norway spruce ([27]; Fig. 1c,d).

24-nt sRNA profiles between pollen and somatic embryonic callus are distinct

In angiosperms, alterations of siRNA size distribution have been observed not only in pollen, but also in cultured cells and DNA methylation-deficient mutants [27, 31–37]. Similarly, cultured cells of Norway spruce also produce 24-nt sRNAs [21]. To investigate whether the TE-derived sRNA profiles are similar between pollen and somatic embryonic callus, we sequenced sRNAs from Norway spruce somatic embryonic callus samples grown at different temperatures: 4 °C, 22 °C, and 28 °C. Consistent with previous work, we also identified a 24-nt peak in our somatic embryonic callus samples ([21]; Fig. 2a, b). The 24-nt sRNAs in the somatic embryonic callus were also exclusively derived from TEs (Fig. 2c, d). Notably, the subtracted ribosomal RNA (rRNA), transfer RNA (tRNA), and small nuclear (nucleolar) RNA (sn(o)RNA)-RNA derived fractions also displayed varying size distributions among tissues (Additional file 1: Figure S1). For rRNA-derived sequences between the 17–42-nt size range, somatic embryonic callus had 4 peaks at 17-nt, 19-nt, 24-nt, and 34-nt, while pollen and needles had peaks at 17-nt, 20-nt and 25-nt. For tRNA-derived sequences, needles and pollen had peaks at 17-nt, compared to a peak at 33-nt in somatic embryonic callus. For sn(o)RNA-related sequences, specifically only somatic embryonic callus had large peaks at 20-nt and 30-nt (Additional file 1: Figure S1). Thus, the r/t/sn(o)RNA processing highly diverged among these tissues.

Temperature was previously shown to affect TE activity in plants [38–40]. For example, high accumulation of TE-derived 21-nt sRNAs was observed after heat shock

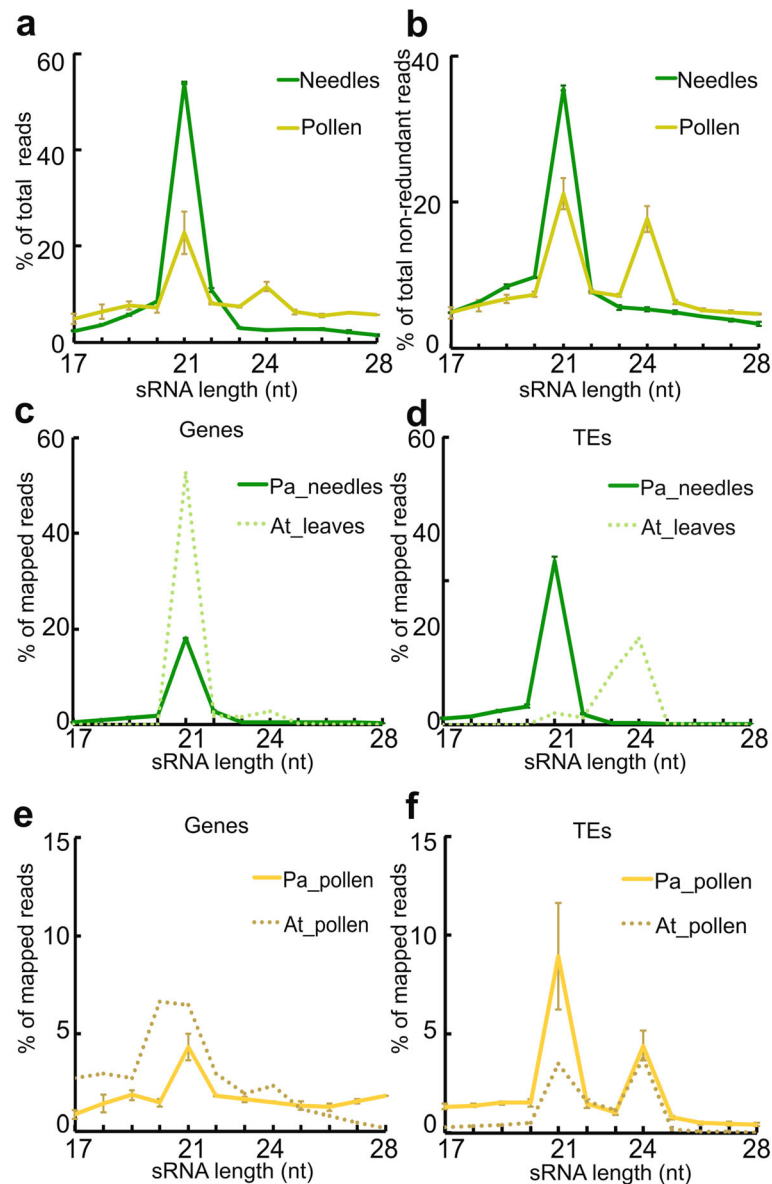
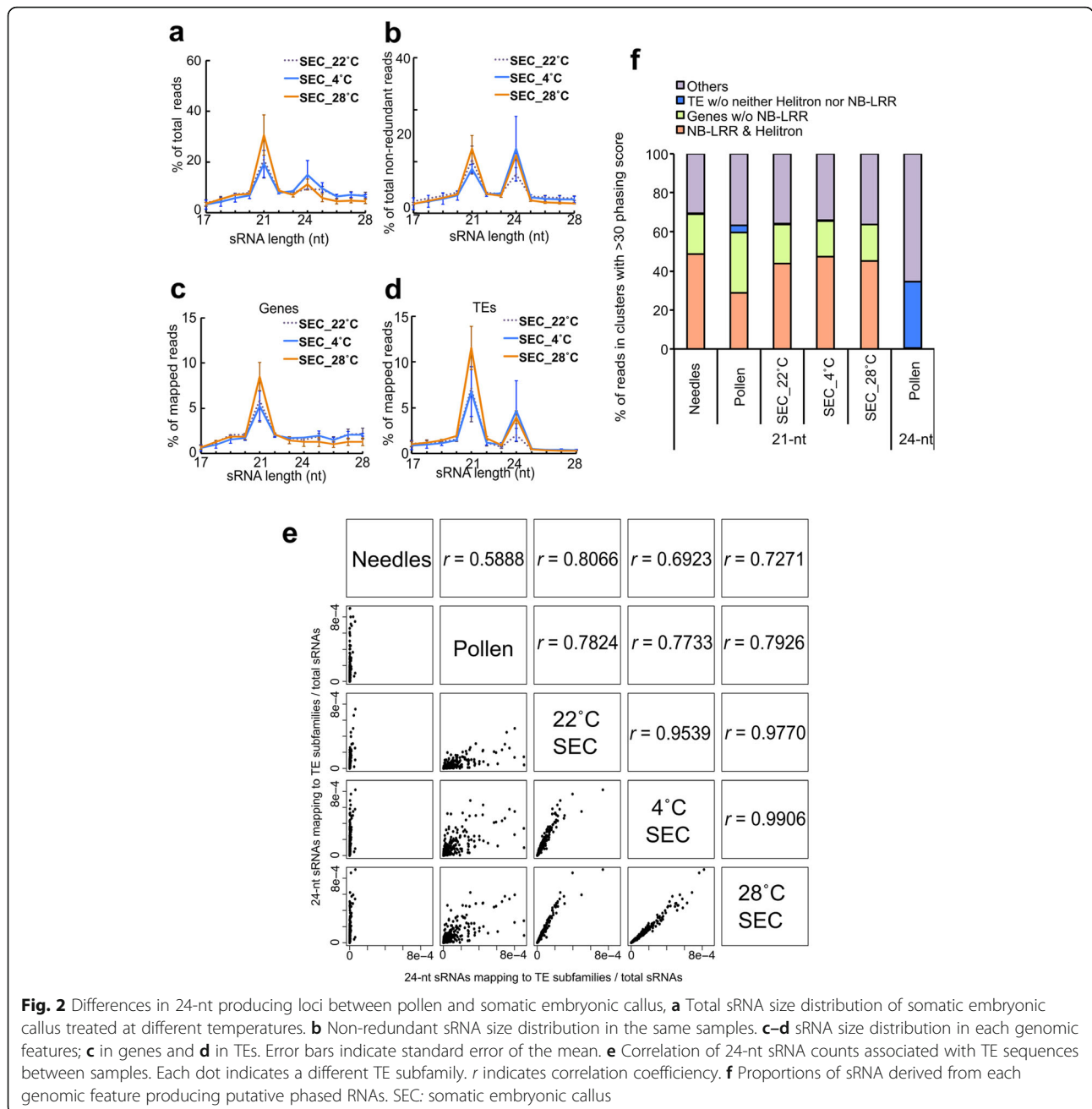


Fig. 1 Size distribution of sRNA in different organs, **a** sRNA size distribution of total read counts in needles and pollen samples after removing transfer RNA (tRNA)-, ribosomal RNA (rRNA)-, small nucleolar RNA (snoRNA)-, and small nuclear (snRNA) RNA-derived sequences. **b** sRNA size distribution of non-duplicated read counts after removing t/r/sn(o)RNA. **c-f** sRNA size distribution; **(c)** in genes, and **(d)** in TEs of *Arabidopsis* and *P.abies* vegetative tissues. **(e)** in genes, and **(f)** in TEs of *Arabidopsis* and *P.abies* pollen. At: *Arabidopsis thaliana*, Pa: *Picea abies*. Error bars indicate standard error of the mean

in *Arabidopsis* [40]. Furthermore, temperature conditions affect miRNA profiles in Norway spruce [41]. Therefore, we compared sRNA profiles of somatic embryonic callus cultured at 4 °C, 22 °C, and 28 °C to address the effect of temperature on sRNA profiles in Norway spruce somatic embryonic callus. TE-associated sRNAs in somatic embryonic callus at all three temperature conditions had obvious peaks at 21-nt and 24-nt length (Fig. 2a-d). To determine whether TE-derived 24-nt sRNAs were generated from the same TE

subfamilies in different tissues and upon temperature stress, we used a detailed annotation of repeat sequences and classified them into different TE subfamilies, and tested for correlation of 24-nt sRNAs mapping to TE subfamilies between different samples. The increase of 24-nt sRNAs occurred equally at different TE subfamilies and was not restricted to specific TE subfamilies (Fig. 2e). Although the temperature conditions varied, the 24-nt sRNA counts correlated well among somatic embryonic callus samples (Fig. 2e). Thus, TE-derived 24-



nt sRNA populations in somatic embryonic callus samples were robust to temperature stress. In contrast, TE-derived 24-nt sRNAs substantially differed between pollen and somatic embryonic callus samples (Fig. 2e), indicating that different TE loci contribute to 24-nt sRNAs in pollen and somatic embryonic callus.

21-nt phasiRNAs contribute a large fraction of the Norway spruce sRNA population [42]. Like in angiosperms [28, 29, 43], also Norway spruce produces reproductive tissue-specific phasiRNAs [42]. Since angiosperm phasiRNA-producing (PHAS) loci produce reproductive-tissue specific 24-nt phasiRNAs, which

so far were not reported in gymnosperms, we tested whether Norway spruce pollen and somatic embryonic callus also generate 24-nt sRNAs in a phased pattern. We searched for phased sRNA clusters using the Shortstack program and extracted sRNA clusters with phasing scores exceeds 30 [44]. As previously reported, we also identified more than 600 phased clusters of 21-nt sRNAs in Norway Spruce needles [42]. In pollen, we identified 160 clusters of phased 21-nt and 10 clusters of phased 24-nt sRNA. In contrast to pollen, almost no phased 24-nt sRNA clusters were found in somatic embryonic callus even when applying

a low threshold (Additional file 1: Figure S2), but there were 180–280 clusters of phased 21-nt sRNA. Previous work revealed that substantial fractions of 21-nt phasiRNA clusters of Norway spruce are derived from nucleotide binding sites and leucine-rich repeat (NBS-LRR) genes [42, 45]. Similarly, we found that approximately 30–50% of 21-nt phasiRNAs in needles and somatic embryonic callus were derived from NBS-LRRs and related sequences (Fig. 2f). The fraction of 21-nt phased RNAs derived from TEs was higher in pollen than in other samples. Most of the 24-nt phased RNA clusters were derived from TEs and not from NB-LRRs and related sequences. So far, we do not know the trigger miRNA for those phased RNAs. Considering the difficulty to precisely identify 24-nt PHAS loci [46], we could not determine whether these 24-nt phased sRNA clusters in pollen are indeed 24-nt PHAS loci. Nevertheless, pollen contains a different type of 24-nt sRNA from somatic embryonic callus, suggesting tissue-specific biogenesis of sRNAs in Norway spruce.

Tissue specific 24-nt sRNAs originate from 21-nt producing loci

In *Arabidopsis*, easiRNA producing loci in pollen and in mutants for the chromatin remodeling factor DDM1 partially overlap with regions that are targeted by 24-nt siRNAs [47]. Furthermore, these tissue-specific 21–22-nt easiRNAs antagonize the production of 24-nt siRNAs due to TE sequences being transcriptionally activated and the transcripts processed by the PTGS pathway [47]. To test whether the same interplay of 24-nt sRNAs and 21-nt sRNAs takes place in Norway spruce, we compared the TE-derived 21-nt and 24-nt sRNA profiles.

Most of TE-derived sRNA populations originated from Gypsy, Copia, and unknown families (Fig. 3a), representing the majority of the genome sequence [18]. Notably, the majority of TE subfamilies generated both 21-nt and 24-nt sRNAs in pollen and somatic embryonic callus and both populations were positively correlated (Fig. 3a, b). Nevertheless, 21-nt fractions were generally more abundant in somatic embryonic callus, while in pollen the 21-nt/ 24-nt ratio in TE subfamilies varied. One of the CMC-EnSpm subfamily and one of the gypsy subfamily produced predominantly 24-nt sRNAs in pollen (Fig. 3a). Consistent with the genome-wide correlation of 21-nt and 24-nt sRNAs, when inspecting individual loci we found similar patterns of 24-nt and 21-nt RNA accumulation (Fig. 3c). Thus, unlike in *Arabidopsis* pollen, 21-nt and 24-nt sRNA accumulated in a parallel manner in pollen and somatic embryonic callus of Norway spruce. The 24-nt-producing loci in pollen exhibited modest 21-nt sRNA accumulation in needles (Fig. 3d), revealing that the same loci generate sRNAs in

pollen and vegetative tissues but are targeted by different pathways.

Expression of RdDM components does not correlate with 24-nt production

DCL3 is required for the accumulation of 24-nt siRNA in angiosperms, and similarly, for the production of 22–24-nt sRNAs in *Physcomitrella patens* [14, 48]. It has been proposed that early land plants did not have distinct subunits for NUCLEAR RNA POLYMERASE D 1 (NRPD1) and NUCLEAR RNA POLYMERASE E 1 (NRPE1), the largest subunits of Pol IV and Pol V, respectively [49]. Consistently, in lycophytes and ferns, 24-nt sRNAs are associated with *NRPE1* expression [49]. However, in the moss *P. patens*, like in angiosperms, NRPD1 is required for 24-nt siRNA production [6, 48], indicating that NRPD1 had a role in 24-nt sRNA production already in Bryophytes. To address the mechanism for tissue-specific sRNA production in Norway spruce, we investigated the expression pattern of homologs of RdDM pathway components and related genes in Norway spruce using the mRNA transcriptome datasets of 22 different tissues/organs/conditions (Congene.org: [18]). We did not find strong correlations between expression of homologs of any RdDM components and 24-nt sRNAs (Additional file 1: Figure S3). Interestingly, among these RdDM homologs, putative homologs of RDR6, MA_10435131g0010 and MA_10435131g0020, which are potentially derived from one gene, correlated in expression with 24-nt sRNAs (Additional file 1: Figure S3), suggesting a possible role of Norway spruce RDR6 in the production of 24-nt sRNAs, similar to what has been proposed in *Arabidopsis* [46].

Discussion

Differences in sRNA populations in gymnosperm tissues

Here, we showed that Norway spruce pollen contains a noticeable amount of 24-nt sRNAs that were exclusively derived from TEs (Fig. 1). Several studies in *Arabidopsis* revealed that 21–22-nt TE-derived sRNAs are generated from transcriptionally active TEs under specific conditions or in specific tissues [27, 47, 50–53]. In contrast, 21-nt TE-derived sRNAs in Norway spruce were prevalent in all examined tissues, including vegetative tissues (Fig. 1), while TE-derived 24-nt sRNAs were tissue-specific. Thus, the pattern of TE-associated sRNA size distributions in Norway spruce is opposite to that in *Arabidopsis*.

Angiosperm genomes also have tissue-specific 24-nt sRNA-producing loci. In *Arabidopsis*, inflorescences and developing siliques have additional loci producing 24-nt siRNA compared to other tissues [14, 29, 54, 55]. Similarly, in rice

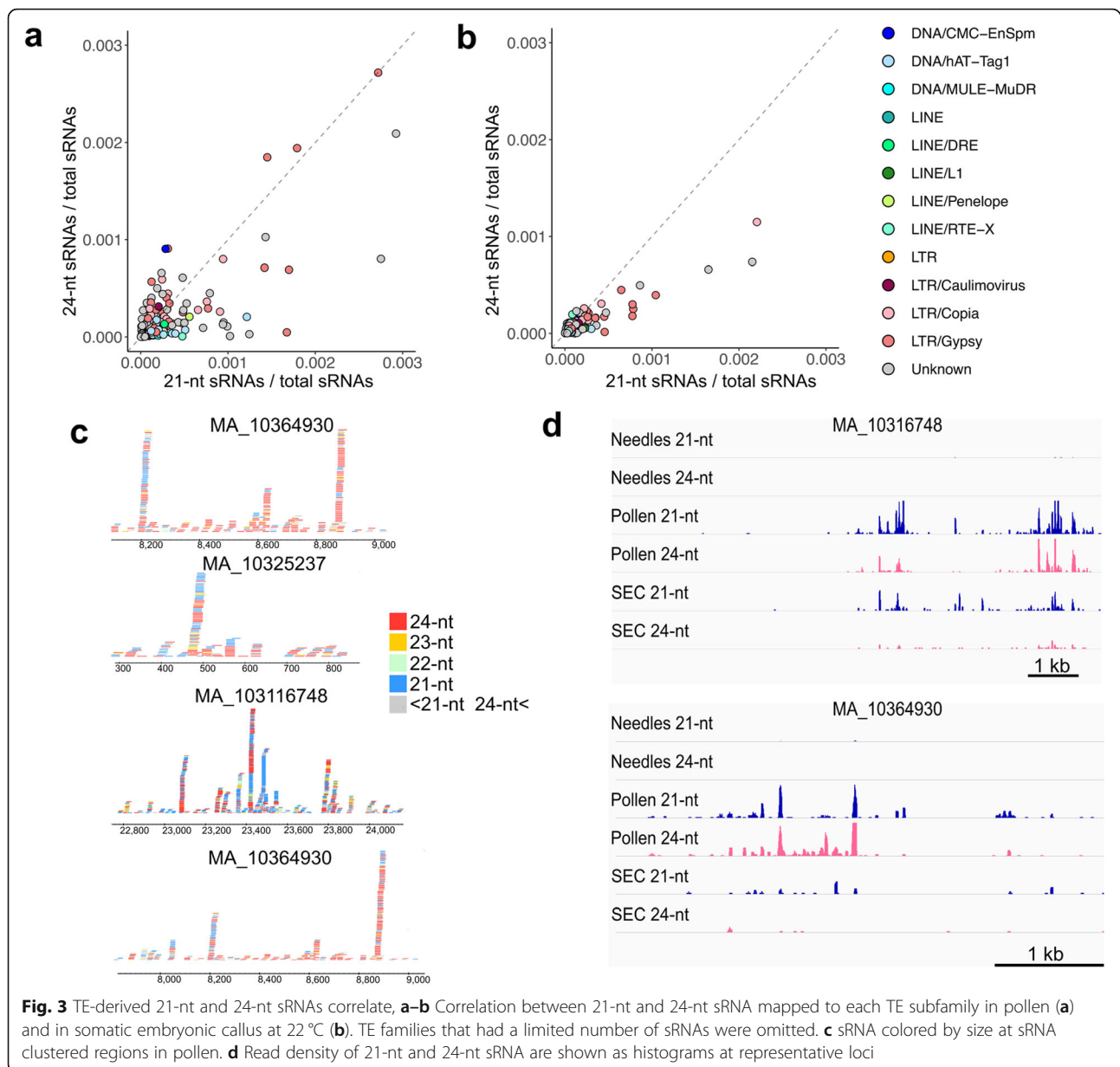


Fig. 3 TE-derived 21-nt and 24-nt sRNAs correlate, **a–b** Correlation between 21-nt and 24-nt sRNA mapped to each TE subfamily in pollen (**a**) and in somatic embryonic callus at 22 °C (**b**). TE families that had a limited number of sRNAs were omitted. **c** sRNA colored by size at sRNA clustered regions in pollen. **d** Read density of 21-nt and 24-nt sRNA are shown as histograms at representative loci

endosperm, there are additional loci generating 24-nt sRNAs [56]. Observations in gymnosperms also revealed divergence of 24-nt sRNA accumulation in different tissues. Developing seeds in *P. glauca* also contain 24-nt sRNA [23]. Similarly, 24-nt sRNAs are present in female flowers of *Pinus tabulaeformis*, but are only a minor fraction in *P. abies* [18, 30]. In some gymnosperms, a substantial fraction of 24-nt sRNAs has been detected even in vegetative tissues [57–60]. Thus, the tissue-specificity of 24-nt sRNA production is diverse in gymnosperms. While Pol IV and Pol V evolved in early land plants, the predominant use of 24-nt-siRNAs seems to have evolved more recently [11, 17, 48, 49]. This may explain the observed differences in 24-nt sRNA production among gymnosperm tissues.

Mature pollen in *Picea* plants consists of a few prothallium cells, a tube cell, and a generative cell, which later on divides into sperm and stalk cells [61, 62]. Except for the existence of the prothallium cells, components of *Picea* pollen are similar to those of angiosperms. It would be interesting to explore which cells contribute the pollen-specific 24-nt sRNA.

In somatic embryonic callus, not only TE-derived sRNAs, but also other sRNAs displayed a tissue-specific size distribution (Additional file 1: Figure S1). Previous work revealed that sRNAs strongly accumulate in *Arabidopsis* cell cultures [63], indicating a conserved requirement of sRNAs during cell proliferation in angiosperms and gymnosperms.

Our results also reveal differences in 24-nt sRNA-producing loci between pollen and somatic embryonic callus (Fig. 2e). In *Arabidopsis*, CLSY chromatin remodeling factors are responsible for locus-specific siRNA production [64]. Two CLSY homologs were differentially expressed among tissues in Norway spruce (Additional file 1: Figure S2), suggesting that similar mechanisms may underlie the tissue-specific production of 24-nt sRNAs in gymnosperms. Furthermore, several putative RdDM components showed higher expression in non-needle tissues, which may connect to the observed tissue-specific accumulation of 24-nt sRNAs (Additional file 1: Figure S2).

Potential roles of 24-nt sRNAs in Norway spruce pollen

In angiosperms, 24-nt sRNAs mediate *de novo* DNA methylation through the RdDM pathway [6]. CHH DNA methylation levels are higher in somatic embryonic callus compared to needles, indicating that 24-nt sRNAs mediate DNA methylation in gymnosperms as well [21]. On the other hand, no obvious increase of DNA methylation has been observed in flower buds [21]. This might reflect the difference in 24-nt sRNA profiles between somatic embryonic callus and pollen reported in our study, although we cannot exclude that 24-nt sRNAs do not induce CHH methylation in pollen, consistent with low CHH methylation levels observed in *Arabidopsis* sperm cells [35, 65].

We identified some loci that produce phased 24-nt RNAs in pollen. Developing pollen in rice and maize contain a large fraction of 21-nt and 24-nt phasiRNAs [28, 29]. The molecular roles of 24-nt phasiRNA are not completely understood, but they are associated with pollen development in monocots [29, 66, 67]. The phased 24-nt loci identified in our analysis do not overlap with previously identified reproductive tissue-specific PHAS loci that produce mainly 21-nt phasiRNA (Additional file 1: Table S2, [42]). These results suggest that gymnosperms have distinct loci producing 21-nt and 24-nt phasiRNAs. In monocots, 24-nt phasiRNAs are associated with meiosis [29, 66] and triggered by miR2275, which the gymnosperm lineage seems to lack [43]. An investigation of meiocytes in Norway spruce might identify more 24-nt phased loci and the trigger miRNA.

Conclusions

Angiosperm pollen has unique sRNA profiles in both dicots and monocots. In this study, we found that pollen of *P.abies* produced 24-nt sRNAs, contrasting the predominant production of 21-nt sRNAs in other tissues. It has been reported that *P.abies* somatic embryonic callus also produces 24-nt sRNAs. Although 24-nt sRNAs are exclusively derived from TEs in both

pollen and somatic embryonic callus, the 24-nt producing loci varied between these tissues. In contrast to *Arabidopsis*, tissue-specific 24-nt sRNAs are not antagonistic and rather positively correlated with 21-nt sRNAs. Our data provide strong evidence for the divergence of TE-derived sRNA processing between angiosperms and gymnosperms. Nevertheless, the specific occurrence of TE-derived 24-nt sRNAs in pollen suggests that epigenetic reprogramming also occurs during gymnosperm pollen development.

Methods

Plant materials and growth conditions

Pollen and needles were harvested from a wild Norway spruce (*Picea abies* (L) Karst.) tree in Uppsala Sweden (latitude: 59°48'47.7"N, longitude: 17°38'20.6"E) between May and August in 2017. The somatic embryonic cell line (11:18:1:1) was established as described [68] and kindly provided by Dr. Sara von Arnold. Somatic embryonic callus was proliferated on half-strength LP [69]. Somatic embryonic callus was grown at 22 °C unless otherwise indicated. Temperature treated callus was exposed to temperature of 4 °C, 22 °C, and 28 °C for 3 weeks in the dark and then immediately frozen in liquid nitrogen.

Library preparation and sequencing of sRNA

Total RNAs were extracted as described previously [70]. Subsequently, sRNAs were separated from high molecular weight RNAs by PEG8000 precipitation according as previously described [71]. sRNA libraries were constructed from sRNA recovered from gels using the NEB-Next® Small RNA Library Prep Set for Illumina (NEB, MA, USA) and subsequently sequenced on an Illumina HiSeq 2500 (50 bp single reads) at the SciLifeLab (Uppsala, Sweden).

Mapping of sRNA sequences

Adaptor sequences were trimmed from read sequences using reaper [72]. r/t/sn(o)RNA sequences were removed using the Rfam database [73]. The 17–28-nt read sequences were mapped to the Pabies1.0-genome-gene-only.fa (<http://congenie.org>) [18] using Bowtie [74], requiring the best match with allowing a maximum two mismatches. The sequencing data of *Arabidopsis* sRNA datasets of *Arabidopsis* were downloaded from Gene Expression Omnibus (GEO) at the National Center for Biotechnology Information (NCBI). The accessions used in this study are as follows; wild-type pollen: GSM1495679, wild-type leaves: GSM1330561. sRNA clusters and phased sRNA clusters were identified by Shortstack [44]. sRNA clusters with > 30 phased score were considered as putative phased sRNA loci.

TE annotation and prediction of *P.abies* homologs

Repeat sequences were annotated using RepeatModeler (Smit, AFA., Hubley, R. RepeatModeler Open-1.0.72008–2015 <<http://www.repeatmasker.org>>) and RepeatMasker (Smit AFA., Hubley R., Green P., RepeatMasker Open-4.0.12013–2015. <<http://www.repeatmasker.org>>) against the *P.abies* genome assembly v1.0 (Pabies1.0-genome.fa) [18]. The numbers of each TE subfamily locus in the Pabies1.0-genome-gene-only.fa correlated well with those in Pabies1.0-genome.fa except for rare TE subfamilies. TE remnants, which are truncated sequences, were also considered as TEs in this study. The rebase that was referred for the TE annotation includes a helitron sequence containing a homolog of an NBS-LRR sequence [75]. Because of a difficulty to distinguish between helitrons and NBS-LRR genes, 3 helitron subfamilies were excluded from this TE annotation. A BLASTP (2.2.30+) search [76, 77] was performed (E value < 0.0001) against the peptide sequence data from *Arabidopsis* TAIR10 gene models to annotate Pabies1.0-all-pep.faa (Congenie).

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12864-019-6385-7>.

Additional file 1: Table S1. sRNA sequencing and mapping data, **Table S2.** 24-nt sRNA clusters with phasing scores with above threshold, **Figure S1.** Size distributions of sRNA derived from ribosomal/transfer/small nuclear (nucleolar) RNA size distributions in needles, pollen, and somatic embryonic callus at 22 °C. SEC: somatic embryonic callus, **Figure S2.** The number of sRNA cluster loci with certain phasing score, **Figure S3.** Expression pattern of RdDM and sRNA related pathway component homologs.

Abbreviations

miRNA: micro RNA (ribonucleic acid); nt: Nucleotides; phasiRNA: Phased siRNA; PTGS: Post transcriptional gene silencing; RdDM: RNA-directed DNA methylation; siRNA: small(short) interfering RNA; sRNA: Small RNA; TE: Transposable element

Acknowledgements

We thank Dr. Sara von Arnold for the technical advice of handling spruce callus. The partial computation works were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at Uppmax (projects b2012015 and snic2017-7-223).

Authors' contributions

MN performed the sample collection and the library preparation. MN, CK, and LH analyzed the data; MN and LH planned the experiments; and MN and CK wrote the manuscript. All authors have read and approved the manuscript.

Funding

This work was supported by Svenska Forskningsrådet Formas (grant no. 2012–1519, 2016–00453 to L.H.) and by Knut och Alice Wallenbergs Stiftelse (grant no. 2012.0087 to L.H.). These funding bodies had no role in data collection, data analysis and interpretation, or in writing the manuscript.

Availability of data and materials

sRNA-Seq data in this study has been deposited in Gene Expression Omnibus NCBI (Accession: GSE129413).

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 26 August 2019 Accepted: 11 December 2019

Published online: 19 December 2019

References

- Willmann MR, Poethig RS. Conservation and evolution of miRNA regulatory programs in plant development. *Curr Opin Plant Biol.* 2007;10(5):503–11.
- Cuperus JT, Fahlgren N, Carrington JC. Evolution and Functional Diversification of MIRNA Genes. *Plant Cell.* 2011;23:431–42.
- Jones-Rhoades MW. Conservation and divergence in plant microRNAs. *Plant Mol Biol.* 2012;80(1):3–16.
- Holoch D, Moazed D. RNA-mediated epigenetic regulation of gene expression. *Nature Reviews Genetics.* 2015;16(2):71–84.
- Siomi MC, Sato K, Pezic D, Aravin AA. PIWI-interacting small RNAs: the vanguard of genome defence. *Nat Rev Mol Cell Biol.* 2011;12(4):246–58.
- Matzke MA, Moshier RA. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat Rev Genet.* 2014;15(6):394–408.
- Gascioli V, Mallory AC, Bartel DP, Vaucheret H. Partially redundant functions of arabidopsis DICER-like enzymes and a role for DCL4 in producing trans-acting siRNAs. *Curr Biol.* 2005;15:1494–500.
- Zhang J, Xu Y, Huan Q, Chong K. Deep sequencing of Brachypodium small RNAs at the global genome level identifies microRNAs involved in cold stress response. *BMC Genomics.* 2009;10:449.
- Li MY, Wang F, Xu ZS, Jiang Q, Ma J, Tan GF, et al. High throughput sequencing of two celery varieties small RNAs identifies microRNAs involved in temperature stress response. *BMC Genomics.* 2014;15:242.
- Zhang Y, Zhu X, Chen X, Song C, Zou Z, Wang Y, et al. Identification and characterization of coldresponsive microRNAs in tea plant (*Camellia sinensis*) and their targets using high-throughput sequencing and degradome analysis. *BMC Plant Biol.* 2014;14:271.
- Ma L, Hatlen A, Kelly LJ, Becher H, Wang W, Kovarik A, et al. Angiosperms are unique among land plant lineages in the occurrence of key genes in the RNA-directed DNA methylation (RdDM) pathway. *Genome Biol Evol.* 2015;7:2648–62.
- Li H, Dong Y, Chang J, He J, Chen H, Liu Q, et al. High-Throughput MicroRNA and mRNA Sequencing Reveals That MicroRNAs May Be Involved in Melatonin-Mediated Cold Tolerance in *Citrus lanatus* L. *Front Plant Sci.* 2016;7:1231.
- Song G, Zhang R, Zhang S, Li Y, Gao J, Han X, et al. Response of microRNAs to cold treatment in the young spikes of common wheat. *BMC Genomics.* 2017;18:212.
- Kasschau KD, Fahlgren N, Chapman EJ, Sullivan CM, Cumbie JS, Givan SA, et al. Genome-wide profiling and analysis of Arabidopsis siRNAs. *PLoS Biol.* 2007;5:e57.
- Zhang X, Henderson IR, Lu C, Green PJ, Jacobsen SE. Role of RNA polymerase IV in plant small RNA metabolism. *Proc Natl Acad Sci.* 2007;104:4536–41.
- El Baidouri M, Kim K Do, Abernathy B, Arikat S, Maumus F, Panaud O, et al. A new approach for annotation of transposable elements using small RNA mapping. *Nucleic Acids Res.* 2015;43:e84.
- Huang Y, Kendall T, Forsythe ES, Dorantes-Acosta A, Li S, Caballero-Pérez J, et al. Ancient origin and recent innovations of RNA polymerase IV and V. *Mol Biol Evol.* 2015;32:1788–99.
- Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin YC, Scofield DG, et al. The Norway spruce genome sequence and conifer genome evolution. *Nature.* 2013;497:579–84.
- Chávez Montes RA, De Fátima Rosas-Cárdenas F, De Paoli E, Accebbi M, Rymarquis LA, Mahalingam G, et al. Sample sequencing of vascular plants demonstrates widespread conservation and divergence of microRNAs. *Nat Commun.* 2014;5:3722.
- Khraiweh B, Arif MA, Seumel GI, Ossowski S, Weigel D, Reski R, et al. Transcriptional Control of Gene Expression by MicroRNAs. *Cell.* 2010;140:111–22.

21. Ausin I, Feng S, Yu C, Liu W, Kuo HY, Jacobsen EL, et al. DNA methylome of the 20-gigabase Norway spruce genome. *Proc Natl Acad Sci*. 2016;113:E8106–13.
22. Zhang J, Wu T, Li L, Han S, Li X, Zhang S, et al. Dynamic expression of small RNA populations in larch (*Larix leptolepis*). *Planta*. 2013;237:89–101.
23. Liu Y, El-Kassaby YA. Landscape of fluid sets of hairpin-derived 21–/24-nt-long small RNAs at seed set uncovers special epigenetic features in *Picea glauca*. *Genome Biol Evol*. 2017;9:82–92.
24. Ujino-Ihara T, Ueno S, Uchiyama K, Futamura N. Comprehensive analysis of small mas expressed in developing male strobili of *Cryptomeria japonica*. *PLoS One*. 2018;13:e0193665.
25. Rodrigues AS, Chaves I, Costa BV, Lin Y-C, Lopes S, Milhinhos A, et al. Small RNA profiling in *Pinus pinaster* reveals the transcriptome of developing seeds and highlights differences between zygotic and somatic embryos. *Sci Rep*. 2019;9:11327.
26. Neale DB, Wegrzyn JL, Stevens KA, Zimin AV, Puiu D, Crepeau MW, et al. Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome Biol*. 2014;15:R59.
27. Slotkin RK, Vaughn M, Borges F, Tanurdžić M, Becker JD, Feijó JA, et al. Epigenetic Reprogramming and Small RNA Silencing of Transposable Elements in Pollen. *Cell*. 2009;136:461–72.
28. Komiya R, Ohyanagi H, Niihama M, Watanabe T, Nakano M, Kurata N, et al. Rice germline-specific Argonaute MEL1 protein binds to phasiRNAs generated from more than 700 lincRNAs. *Plant J*. 2014;78:385–97.
29. Zhai J, Zhang H, Arikrit S, Huang K, Nan G-L, Walbot V, et al. Spatiotemporally dynamic, cell-type-dependent premeiotic and meiotic phasiRNAs in maize anthers. *Proc Natl Acad Sci*. 2015;112:3146–51.
30. Niu SH, Liu C, Yuan HW, Li P, Li Y, Li W. Identification and expression profiles of sRNAs and their biogenesis and action-related genes in male and female cones of *Pinus tabulaeformis*. *BMC Genomics*. 2015;16:693.
31. Kaepler SM, Phillips RL. Tissue culture-induced DNA methylation variation in maize. *Proc Natl Acad Sci*. 1993;90:8773–6.
32. Law RD, Suttle JC. Chromatin remodeling in plant cell culture: patterns of DNA methylation and histone H3 and H4 acetylation vary during growth of asynchronous potato cell suspensions. *Plant Physiol Biochem*. 2005;43:527–34.
33. Tanurdzic M, Vaughn MW, Jiang H, Lee TJ, Slotkin RK, Sosinski B, et al. Epigenomic consequences of immortalized plant cell suspension culture. *PLoS Biol*. 2008;6:e2880–95.
34. Neelakandan AK, Wang K. Recent progress in the understanding of tissue culture-induced genome level changes in plants and potential applications. *Plant Cell Reports*. 2012;31(4):597–620.
35. Calarco JP, Borges F, Donoghue MTA, Van Ex F, Jullien PE, Lopes T, et al. Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. *Cell*. 2012;151:194–205.
36. Vining K, Pomraning KR, Wilhelm LJ, Ma C, Pellegrini M, Di Y, et al. Methylome reorganization during in vitro dedifferentiation and regeneration of *Populus trichocarpa*. *BMC Plant Biol*. 2013;13:92.
37. Walker J, Gao H, Zhang J, Aldridge B, Vickers M, Higgins JD, et al. Sexual-lineage-specific DNA methylation regulates meiosis in *Arabidopsis*. *Nat Genet*. 2018;50:130–7.
38. Carpenter R, Martin C, Coen ES. Comparison of genetic behaviour of the transposable element Tam3 at two unlinked pigment loci in *Antirrhinum majus*. *MGG Mol Gen Genet*. 1987;207:82–9.
39. Miller SM, Kirk DL, Schmitt R. Jordan, an active volvox transposable element similar to higher plant transposons. *Plant Cell*. 1993;5:1125–38.
40. Ito H, Gaubert H, Bucher E, Mirouze M, Vaillant I, Paszkowski J. An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature*. 2011;472:115–9.
41. Yakovlev IA, Fossdal CG, Johnsen Ø. MicroRNAs, the epigenetic memory and climatic adaptation in Norway spruce. *New Phytol*. 2010;187:1154–69.
42. Xia R, Xu J, Arikrit S, Meyers BC. Extensive families of miRNAs and PHAS loci in Norway spruce demonstrate the origins of complex phasiRNA networks in seed plants. *Mol Biol Evol*. 2015;32:2905–18.
43. Xia R, Chen C, Pokhrel S, Ma W, Huang K, Patel P, et al. 24-nt reproductive phasiRNAs are broadly present in angiosperms. *Nat Commun*. 2019;10:627.
44. Axtell MJ. ShortStack: comprehensive annotation and quantification of small RNA genes. *RNA*. 2013;19:740–51.
45. Källman T, Chen J, Gyllenstrand N, Lagercrantz U. A significant fraction of 21-nucleotide small RNA originates from phased degradation of resistance genes in several perennial species. *Plant Physiol*. 2013;162:741–54.
46. Polydore S, Lunardon A, Axtell MJ. Several phased siRNA annotation methods can frequently misidentify 24 nucleotide siRNA-dominated PHAS loci. *Plant Direct*. 2018;2:e00101.
47. Creasey KM, Zhai J, Borges F, Van Ex F, Regulski M, Meyers BC, et al. MiRNAs trigger widespread epigenetically activated siRNAs from transposons in *Arabidopsis*. *Nature*. 2014;508:411–5.
48. Sung HC, Addo-Quaye C, Coruh C, Arif MA, Ma Z, Frank W, et al. Physcomitrella patens DCL3 is required for 22-24 nt siRNA accumulation, suppression of retrotransposon-derived transcripts, and normal development. *PLoS Genet*. 2008;4:e1000314.
49. You C, Cui J, Wang H, Qi X, Kuo LY, Ma H, et al. Conservation and divergence of small RNA pathways and microRNAs in land plants. *Genome Biol*. 2017;18:158.
50. McCue AD, Nuthikattu S, Reeder SH, Slotkin RK. Gene expression and stress response mediated by the epigenetic regulation of a transposable element small RNA. *PLoS Genet*. 2012;8:e1002474.
51. Pontier D, Picart C, Roudier F, Garcia D, Lahmy S, Azevedo J, et al. NERD, a plant-specific GW protein, Defines an Additional RNAi-Dependent Chromatin-Based Pathway in *Arabidopsis*. *Mol Cell*. 2012;48:121–32.
52. Nuthikattu S, McCue AD, Panda K, Fultz D, DeFraia C, Thomas EN, et al. The initiation of epigenetic silencing of active transposable elements is triggered by RDR6 and 21-22 nucleotide small interfering RNAs. *Plant Physiol*. 2013;162:116–31.
53. Marí-Ordóñez A, Marchais A, Etcheverry M, Martin A, Colot V, Voinnet O. Reconstructing de novo silencing of an active plant retrotransposon. *Nat Genet*. 2013;45:1029–39.
54. Lu C, Tej SS, Luo S, Haudenschild CD, Meyers BC, Green PJ. Genetics: Elucidation of the small RNA component of the transcriptome. *Science (80-)*. 2005;309:1567–9.
55. Swiezewski S, Crevillen P, Liu F, Ecker JR, Jerzmanowski A, Dean C. Small RNA-mediated chromatin silencing directed to the 3' region of the *Arabidopsis* gene encoding the developmental regulator, FLC. *Proc Natl Acad Sci*. 2007;104:3633–8.
56. Rodrigues JA, Ruan R, Nishimura T, Sharma MK, Sharma R, Ronald PC, et al. Imprinted expression of genes and small RNA is associated with localized hypomethylation of the maternal genome in rice endosperm. *Proc Natl Acad Sci U S A*. 2013;110:7934–9.
57. Wan LC, Wang F, Guo X, Lu S, Qiu Z, Zhao Y, et al. Identification and characterization of small non-coding RNAs from Chinese fir by high throughput sequencing. *BMC Plant Biol*. 2012;12:146.
58. Hao DC, Yang L, Xiao PG, Liu M. Identification of *Taxus* microRNAs and their targets with high-throughput sequencing and degradome analysis. *Physiol Plant*. 2012;146:388–403.
59. Zhang Q, Li J, Sang Y, Xing S, Wu Q, Liu X. Identification and characterization of MicroRNAs in *Ginkgo biloba* var. *epiphylla* mak. *PLoS One*. 2015;10:e0127184.
60. Galdino JH, Eguiluz M, Guzman F, Margis R. Novel and conserved miRNAs among Brazilian pine and other gymnosperms. *Front Genet*. 2019;10:222.
61. Hejnowicz A. Anatomy, embryology, and karyology. *Bud Structure Shoot Development*. 2007.
62. Fernando DD, Quinn CR, Brenner ED, Owens JN. Male Gametophyte Development and Evolution in Extant Gymnosperms. *Int J Plant Dev Biol*. 2010; 4 special issue 1:47–63.
63. Ohtani M, Sugiyama M. Involvement of SRD2-mediated activation of snRNA transcription in the control of cell proliferation competence in *Arabidopsis*. *Plant J*. 2005;43:479–90.
64. Zhou M, Palanca AMS, Law JA. Locus-specific control of the de novo DNA methylation pathway in *Arabidopsis* by the CLASSY family. *Nat Genet*. 2018; 50:865–73.
65. CA I, X F, VK S, TF H, R U, JA R, et al. Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. *Science (80-)*. 2012;337:1360–4.
66. Fei Q, Yang L, Liang W, Zhang D, Meyers BC. Dynamic changes of small RNAs in rice spikelet development reveal specialized reproductive phasiRNA pathways. *J Exp Bot*. 2016;67:6037–49.
67. Ono S, Liu H, Tsuda K, Fukai E, Tanaka K, Sasaki T, et al. EAT1 transcription factor, a non-cell-autonomous regulator of pollen production, activates meiotic small RNA biogenesis in rice anther tapetum. *PLoS Genet*. 2018;14:e1007238.
68. von Arnold S, Clapham D. Spruce embryogenesis. *Methods Mol Biol*. 2008; 427:31–47.
69. Von Arnold S, Eriksson T. In vitro studies of adventitious shoot formation in *Pinus contorta*. *Can J Bot*. 1981;59:870–4.

70. Gambino G, Perrone I, Gribaudo I. A rapid and effective method for RNA extraction from different tissues of grapevine and other woody plants. *Phytochem Anal.* 2008;19:520–5.
71. Peng J, Xia Z, Chen L, Shi M, Pu J, Guo J, et al. Rapid and efficient isolation of high-quality small RNAs from recalcitrant plant species rich in polyphenols and polysaccharides. *PLoS One.* 2014;9:e95687.
72. Davis MPA, van Dongen S, Abreu-Goodger C, Bartonicek N, Enright AJ. Kraken: A set of tools for quality control and analysis of high-throughput sequence data. *Methods.* 2013;63:41–9.
73. Kalvari I, Argasinska J, Quinones-Olvera N, Nawrocki EP, Rivas E, Eddy SR, et al. Rfam 13.0: Shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res.* 2018;46:D335–42.
74. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 2009;10:R25.
75. Bao W, Kojima KK, Kohany O. Repbase update, a database of repetitive elements in eukaryotic genomes. *Mob DNA.* 2015;6:11.
76. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research.* 1997;25(17):3389–402.
77. Schmid MW, Giraldo-Fonseca A, Rövekamp M, Smetanin D, Bowman JL, Grossniklaus U. Extensive epigenetic reprogramming during the life cycle of *Marchantia polymorpha*. *Genome Biol.* 2018;19:9.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

