

Sequencing of small RNAs of the fern *Pleopeltis minima* (Polypodiaceae) offers insight into the evolution of the microRNA repertoire in land plants

Florencia Berruezo¹, Flávio S. J. de Souza^{3,4}, Pablo I. Picca⁵, Sergio I. Nemirovsky⁶, Leandro Martínez-Tosar^{1,2}, Mercedes Rivero⁷, Alejandro N. Mentaberry^{1,2}, Alicia M. Zelada^{1,2*}

¹Laboratorio de Agrobiotecnología, Departamento de Fisiología, Biología Molecular y Celular, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina

²Instituto de Biodiversidad y Biología Experimental y Aplicada, Consejo Nacional de Investigaciones Científicas y Técnicas-Universidad de Buenos Aires (IBBEA, CONICET-UBA), Buenos Aires, Argentina

³Instituto de Investigaciones en Ingeniería Genética y Biología Molecular "Dr Héctor N. Torres" (INGEBI-CONICET), Buenos Aires, Argentina

⁴Departamento de Fisiología, Biología Molecular y Celular, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina

⁵Departamento de Biodiversidad y Biología Experimental, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina

⁶Instituto de Química Biológica de la Facultad de Ciencias Exactas y Naturales (IQUIBICEN, CONICET-UBA), Universidad de Buenos Aires, Buenos Aires, Argentina

⁷Instituto de Agrobiotecnología de Rosario (INDEAR), Rosario, Santa Fe, Argentina.

*Author for correspondence: azelada@fbmc.fcen.uba.ar

Abstract

MicroRNAs (miRNAs) are short, single stranded RNA molecules that regulate the stability and translation of messenger RNAs in diverse eukaryotic groups. Several miRNA genes are of ancient origin and have been maintained in the genomes of animal and plant taxa for hundreds of millions of years, and functional studies indicate that ancient miRNAs play key roles in development and physiology. In the last decade, genome and small RNA (sRNA) sequencing of several plant species have helped unveil the evolutionary history of land plant miRNAs. Land plants are divided into bryophytes (liverworts, mosses), lycopods (clubmosses and spikemosses), monilophytes (ferns and horsetails), gymnosperms (cycads, conifers and allies) and angiosperms (flowering plants). Among these, the fern group occupies a key phylogenetic position, since it represents the closest extant cousin taxon of seed plants, i.e. gymno- and angiosperms. However, in spite of their evolutionary, economic and ecological importance, no fern genome has been sequenced yet and few genomic resources are available for this group. Here, we sequenced the small RNA fraction of an epiphytic South American fern, *Pleopeltis minima* (Polypodiaceae), and compared it to plant miRNA databases, allowing for the identification of miRNA families that are shared by all land plants, shared by all vascular plants (tracheophytes) or shared by euphyllophytes (ferns and seed plants) only. Using the recently described transcriptome of another fern, *Lygodium japonicum*, we also estimated the degree of conservation of fern miRNA targets in relation to other plant groups. Our results pinpoint the origin of several miRNA families in the land plant evolutionary tree with more precision and are a resource for future genomic and functional studies of fern miRNAs.

keywords: *Polypodium squalidum*, microRNAs, evolution, fern genomics

Introduction

Land plants (embryophytes) evolved from fresh water, streptophyte green algae during the Ordovician period, over 470 million years ago (MYA) [1, 2]. Major diversification events between the Silurian and Permian periods (385-470 MYA) gave rise to groups represented today by bryophytes *sensu lato* (liverworts, hornworts and mosses), lycophytes (clubmosses and spikemosses) and the euphyllophytes, comprising monilophytes (ferns) and spermatophytes (seed plants) [1]. Extant seed plants are divided into gymnosperms (cycads, conifers) and angiosperms. Early diversification in the Palaeozoic resulted in different strategies of alternating (gametophytic and sporophytic) generations and the origination of key anatomical features in different groups of plants, including vascular systems and various types of meristems, leaves and roots, as well as seeds [1, 3]. The origin and evolution of these and other features is studied with fossils and comparative anatomical and physiological studies, but comparative genomics of extant species can also help understand plant evolution. In the last decade, several angiosperm genomes have been sequenced, as well as the genomes of a moss (*Physcomitrella patens*), a lycopod (the spikemoss *Selaginella moellendorffii*) and a conifer tree (Norway spruce, *Picea abies*) [4-6]. Genomic analyses show that land plants share a common set of transcriptional regulators, including transcription factors and microRNAs, that set them aside from green algae [3, 7].

MiRNA genes are transcribed as long, primary RNA precursors (pri-miRNAs) that can form a foldback, hairpin structure. In plants, pri-mRNAs are processed by Dicer-like 1 (DCL1) protein to generate a miRNA/miRNA* duplex. The miRNA* ("star") strand is generally degraded, whereas the mature miRNA molecule is incorporated into a RNA-induced silencing complex (RISC) that contains Argonaute 1 protein (AGO1). The ~21nucleotide (nt) miRNA associated with RISC serves as a guide to recognize specific messenger RNAs (mRNAs), interfering with their translation or leading to mRNA degradation [8, 9]. MiRNAs are, thus, negative regulators of gene expression. Analyses of land plant genomes and small-RNA transcriptomics have revealed that many miRNA genes are common to all plant groups, suggesting that they originated before or during the colonization of land by plant ancestors. Many of these deeply conserved miRNAs play

important roles in plant development and have likely contributed to the diversification of land plant body plans during evolution [7, 10-13].

Our understanding of plant miRNA evolution is hampered by biased sampling, since most genomic and transcriptomic studies have been carried out in angiosperms. In particular, there is a dearth of knowledge on the miRNA repertoire of ferns. The fern clade (Monilophyta) [14] comprises today around 12,000 species. It occupies a special place in the plant phylogenetic tree as the sister group of seed plants (Fig 1A). Thus, the study of ferns should illuminate critical aspects of plant evolution, like the transition from homosporous to strict heterosporous, sporophyte dominance, as well as aspects of the development of leaves, roots and vascular systems that distinguish ferns and seed plants [3, 15]. No fern nuclear genome has been sequenced to date, although some genomic surveys and transcriptomic analyses have been carried out [16, 17]. An early survey based on hybridization arrays indicated the presence of eight miRNAs in the fern *Ceratopteris thalictroides* that are also conserved in other land plants [18]. More recently, a broad survey of several land plants included an aquatic fern, *Marsilea quadrifolia*, which added some conserved miRNAs to the fern repertoire [19].

Here, we study a South American fern, *Pleopeltis minima* (Bory) J. Prado & R.Y. Hirai (synonyms: *Polypodium squalidum* Vell.; *Pleopeltis squalida* (Vell.) de la Sota) [20]. *P. minima* is a small, mostly epiphytic fern belonging to the Polypodiaceae family that occurs in forests in Argentina, Bolivia, Southern Brazil, Paraguay and Uruguay. Its scaly rhizome bears several 2-10 cm long compound leaves (fronds), divided into pinnae (leaflets). Fertile leaves possess exindusiate sori (cluster of sporangia) of round shape on the abaxial side (Fig 1B). One particular characteristic of *P. minima* is its dissection tolerance (poikilohydry) [21, 22]. We have sequenced the small RNA fraction of *P. minima* and searched for miRNAs phylogenetically conserved in other plants. We present a group of conserved fern miRNAs and their predicted targets, thereby advancing our understanding of miRNA evolution in the fern and land plant lineage.

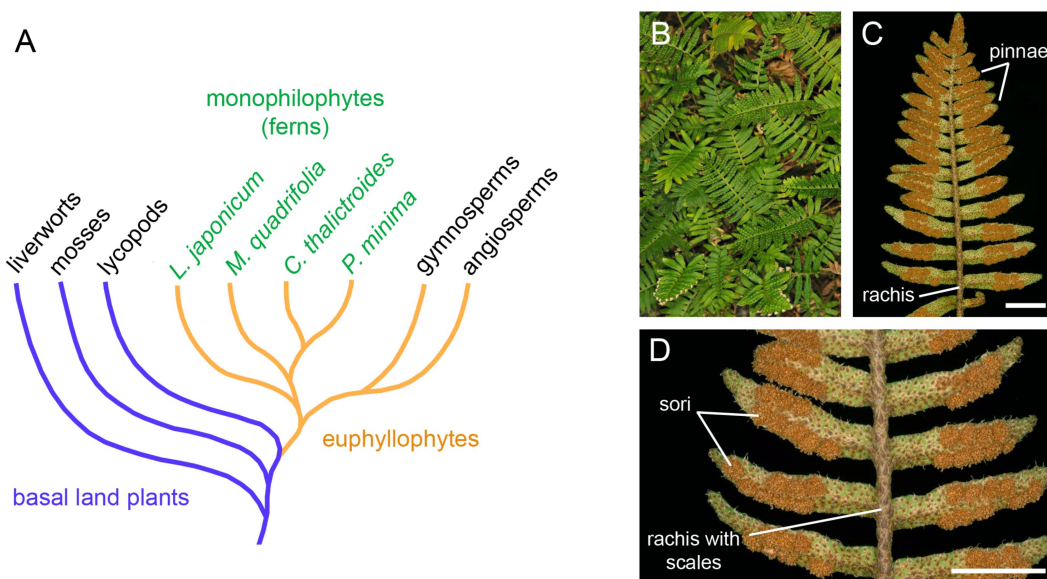


Fig 1. Phylogenetic position and habit of *Pleopeltis minima*. (A) Simplified summary of phylogenetic relationships among major plant lineages relevant for the present study [3, 14, 29]. The fern (“monilophytes”) clade is the sister group to seed plants, gymnosperms and angiosperms. Within ferns, *P. minima* belongs to the order Polypodiales and is more closely related to *Ceratopteris thalictroides*, while *Marsilea quadrifolia* belongs to the order Salviniales and *Lygodium japonicum* to the order Schizaeales. (B) Habit of *P. minima* growing as epiphyte upon a tree trunk. (C) Abaxial side of leaf surface showing form and organization. (D) Portion of fertile leaf showing position of sori and scales. Scale bars: 5 mm.

Results

Sequencing and analysis of *P. minimas* RNAs

Small RNAs (sRNAs) were extracted from two samples of fertile fronds of *P. minima*, converted to cDNA and subjected to Illumina sequencing, generating a total of 25,947,729 reads, of which 20,989,537 were high-quality reads with lengths between 15 and 45 nt. Of these, 6,268,973 reads corresponding to mRNA fragments, tRNAs, rRNAs,

snoRNAs, snRNAs and repetitive elements were discarded, leaving a total of 14,720,564 sRNA reads to be further analysed (S1 Table).

The length of functional sRNAs are thought to lay in the 20-24 nucleotide (nt) range. In angiosperms, sRNA length distribution typically displays two peaks at 21 and 24 nt, with miRNAs centred around 21 nt, while 24 nt sRNAs are associated with silencing of retroposons by heterochromatinization [9]. The lycopod *S. moellendorffii* and the spruce *Picea abies* both lack a prominent 24 nt sRNA peak [5, 6]. In *P. minima*, however, we observed a prominent 21 nt peak and a smaller, but noticeable, peak at 24 nt (Fig 2). A similar distribution is found in the aquatic fern *M. quadrifolia* [19], showing that the formation of sRNAs of both classes is a common characteristic of ferns.

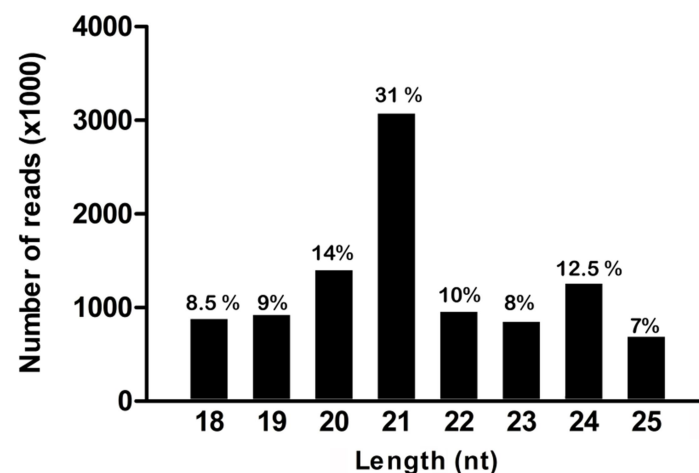


Fig 2. Size distribution of *P. minima* sRNA sequences. The graph displays the raw number of sRNA reads in the range of 18 to 25 nt. The percentage of reads in each size category in the range is indicated. A large peak at 21 nt and a smaller peak at 24 nt are noticeable.

Identification of conserved *P. minima* miRNAs

Since *P. minima* lacks a genome sequence and no fern genome is available for comparison, we focused on finding miRNAs in our samples that are conserved in other land plant genomes. To this end, the 16,482,278 sRNA reads of *P. minima* were compared to

version 21 of miRBase [23], yielding 1,127,970 reads that could be aligned to miRNAs or MIRNA genes in the database, corresponding to 3,912 unique sequences (S1 Table).

It is known that the confidence on the existence of miRNA families in miRBase varies widely, and the evidence for many miRNA families is rather weak, specially for miRNAs coming from species lacking a genome sequence [8]. In view of this, to identify bona fide *P. minima* miRNAs we required the miRNA families 1) to be represented by at least 10 reads of 20-22 nt in our data; 2) to exhibit consistent 5' cleavage processing of the mature miRNA strand and 3) to be identical or very similar to highly confident sequences of mature miRNA or miRNA* strands of other plant species, as defined by [8]. For some miRNA families, namely miR162, miR395 and miR477, even though the total number of reads was under 10, we still considered them to be real miRNAs since the reads included both miRNA and miRNA* strands that were identical or nearly identical in sequence to the corresponding miRNA and miRNA* sequences from other plants (S1 Fig).

The analysis revealed a total of 57 conserved miRNAs in *P. minima*, belonging to 23 miRNA families (Table 1). Since miR156 is related in sequence and function to miR529 [24, 25] and the same occurs between miR159 and miR319 [26, 27], the number of conserved *P. minima* miRNA families could alternatively be considered as being 21. For five families (pmi-miR162, pmi-miR390, pmi-miR395, pmi-miR477 and pmi-miR529), sequences that represent the putative miRNA/miRNA* duplex strands (i.e., the 5p and 3p strands of the miRNAs) were identified (S1 Fig). Sixty-one percent (35/57) of the conserved *P. minima* miRNAs, belonging to 18 of the 23 miRNA families, have a 5' terminal uridine residue (Table 1), a conserved feature of miRNAs recognized by the AGO1 protein, while miRNAs from the pmi-miR390, pmi-477 and pmi-miR529 families have mostly a 5' terminal adenine residue, a feature found in miRNAs recognized preferentially by AGO2 and AGO4 in angiosperms [28].

Table 1. Conserved *P. minima* miRNAs.

| miRNA family | miRNA sequence | Given miRNA name | Length | Total reads | Reference miRNA | E-value |
|--------------|------------------------|------------------|--------|-------------|-----------------|---------|
| miR156 | TTGACAGAAGATAGAGAGCAC | pmi-miR156_v1 | 21 nt | 1898 | ath-miR157a 5p | 0.002 |
| | TGACAGAAGATAGAGAGCAC | pmi-miR156_v2 | 20 nt | 225 | ath-miR157d 5p | 0.004 |
| | TGACAGAAGATAGAGAGCACTT | pmi-miR156_v3 | 22nt | 14 | ath-miR157a 5p | 0.004 |
| miR159 | TTTGGATTGAAGGGAGCTCTA | pmi-miR159_v1 | 21 nt | 38 | ath-miR159a 3p | 0.002 |
| | TTTGGATTGAAGGGAGCTCC | pmi-miR159_v2 | 20 nt | 29 | ath-miR159c 3p | 0.004 |
| miR160 | TTGCCTGGCTCCCTGCATGCC | pmi-miR160_v1 | 21 nt | 23027 | nta-miR160d 5p | 0.004 |
| | TTGCCTGGCTCCCTGCATGCA | pmi-miR160_v2 | 21nt | 353 | nta-miR160d 5p | 0.004 |
| | TGCCTGGCTCCCTGCATGCCA | pmi-miR160_v3 | 21nt | 278 | nta-miR160d 5p | 0.002 |
| miR162 | GGAGGCAGCGGTTTCATCGATC | pmi-miR162_v1 | 21 nt | 2 | aly-miR162a 5p | 0.002 |
| | TCGATAAACCTCTGCATCCAG | pmi-miR162_v2 | 21 nt | 2 | aly-miR162a 3p | 0.002 |
| miR166 | TCGGACCAGGCTTCATTCCCC | pmi-miR166_v1 | 21 nt | 290254 | aly-miR166a 3p | 0.002 |
| | TCGGACCAGGCTTCATTCCCT | pmi-miR166_v2 | 21 nt | 16666 | osa-miR166m 3p | 0.002 |
| | TCGGACCAGGCTTCATTCCCA | pmi-miR166_v3 | 21 nt | 1187 | aly-miR166b 3p | 0.004 |
| | TCGGACCAGGCTTCATTCTT | pmi-miR166_v4 | 20 nt | 298 | cme-miR166i 3p | 0.011 |
| | TCGGACCAGGCTTCATCCCCC | pmi-miR166_v5 | 21 nt | 186 | ath-miR165a 3p | 0.002 |
| | TCGGACCAGGCTTCATTCCCCC | pmi-miR166_v6 | 22nt | 53 | hbr-miR166b 3p | 6e-10-4 |
| | TCGGACCAGGCTTCATTGCCT | pmi-miR166_v7 | 21 nt | 20 | aly-miR166d 3p | 0.008 |
| miR167 | TGAAGCTGCCAGCATGATCTGA | pmi-miR167_v1 | 22 nt | 6 | bdi-miR167d 5p | 6e-10-4 |
| | TGAAGCTGCCAGCATGATCTGG | pmi-miR167_v2 | 22 nt | 3 | ath-miR167d 5p | 6e-10-4 |
| | TGAAGCTGCCAGCATGATCTGC | pmi-miR167_v3 | 22 nt | 2 | ata-miR167c 5p | 6e-10-4 |
| miR168 | TCGCTTGGTGCAGGTCGGGAA | pmi-miR168_v1 | 21 nt | 133 | ath-miR168a 5p | 0.002 |
| | TCGATTGGTGCAGATCGGGA | pmi-miR168_v2 | 20 nt | 15 | osa-miR168a 5p | 0.024 |
| miR169 | GGCAAGTGGTCTTGGCTACCT | pmi-miR169_v2 | 22nt | 25 | ath-miR169a 3p | 0.024 |
| | CGGCAAGTTGTCCTCGGCTAC | pmi-miR169_v1 | 21 nt | 19 | ath-miR169a 3p | 0.024 |
| miR171 | TTGAGCCGTGCCAATATCACAT | pmi-miR171_v1 | 22 nt | 440 | smo-miR171b 3p | 0.002 |
| miR172 | CGAGAATCTTGATGATGCTGC | pmi-miR172_v1 | 21 nt | 653 | vvi-miR172d 3p | 0.004 |
| | CGAGAATCCTGATGATGCTGC | pmi-miR172_v2 | 21 nt | 335 | mtr-miR172a 3p | 0.01 |
| | TGAGAATCTTGATGATGCTGC | pmi-miR172_v3 | 21 nt | 64 | vvi-miR172d 3p | 0.002 |
| | GGAATCTTGATGATGCTGCAA | pmi-miR172_v4 | 21 nt | 19 | ath-miR172e 3p | 0.004 |
| | TGCAGCACCATCAAGATTAC | pmi-miR172_v5 | 21 nt | 25 | aly-miR172b 5p | 0.004 |
| miR319 | CTTGACTGAAGGGAGCTCCC | pmi-miR319_v1 | 21 nt | 1972 | ppt-miR319d 3p | 0.002 |
| | TTGGACTGAAGGGAGCTCCT | pmi-miR319_v2 | 20 nt | 1445 | ath-miR319c 3p | 0.004 |
| | CTTGACTGAAGGGAGCTCCT | pmi-miR319_v3 | 21 nt | 113 | mes-miR319h 3p | 0.002 |
| miR390 | AAGCTCAGGAGGATAGCGCC | pmi-miR390_v1 | 21 nt | 9170 | ath-miR390a 5p | 0.002 |
| | AAGCTCAGGAGGATAGCGCCT | pmi-miR390_v2 | 22 nt | 489 | ath-miR390a 5p | 0.002 |
| | AAGCTCAGGAGGATAGCGCTT | pmi-miR390_v3 | 22 nt | 15 | ath-miR390a 5p | 0.004 |
| | CGCTATCTATCCTGAGTTTCA | pmi-miR390_v4 | 21 nt | 1 | ath-miR390a 3p | 0.008 |
| miR395 | CTGAAGTGTTTGGGGAACTCT | pmi-miR395_v1 | 22 nt | 2 | aly-miR395e 3p | 0.002 |
| | CTGAAGTGTTTGGGGAACTC | pmi-miR395_v2 | 21 nt | 1 | aly-miR395e 3p | 0.002 |
| | GTTCCTCTGAACACTTCATT | pmi-miR395_v3 | 20 nt | 1 | aly-miR395e 5p | 0.024 |
| miR396 | CTCAAGAAAGCTGTGGGAAAA | pmi-miR396_v1 | 21 nt | 414 | ath-miR396b 3p | 0.004 |
| | TTCCACAGCTTTCTTGAACGT | pmi-miR396_v2 | 21 nt | 145 | ath-miR396a 5p | 0.002 |
| miR403 | TTAGATTACAGCACAAACTCG | pmi-miR403_v1 | 21 nt | 16 | ath-miR403 3p | 0.002 |
| miR408 | TGCACTGCCTCTTCCCTGGCT | pmi-miR408_v1 | 21 nt | 88 | smo-miR408 3p | 0.002 |
| | TGCACTGCCTCTTCCCTGGCTC | pmi-miR408_v2 | 22 nt | 51 | smo-miR408 3p | 0.002 |
| miR477 | AGAAGCCTTTGGGGAGAGGG | pmi-miR477_v1 | 21 nt | 5 | ppt-miR477a 3p | 0.022 |
| | CTCTCCCTCAAAGGCTTCCA | pmi-miR477_v2 | 20 nt | 3 | ppt-miR477a 5p | 0.004 |
| miR529 | AGAAGAGAGAGAGCAGCCCC | pmi-miR529_v1 | 21 nt | 289 | ppt-miR529d 5p | 0.002 |
| | AGAAGAGGAGAGCAGCCCC | pmi-miR529_v2 | 21 nt | 256 | ppt-miR529d 5p | 0.008 |
| | AGAAGAGAGAGAGCAGCCT | pmi-miR529_v3 | 21 nt | 23 | bdi-miR529 5p | 0.004 |
| | GCTGTGCTCACTCTCTTCTGG | pmi-miR529_v4 | 21 nt | 163 | ppt-miR529d 3p | 0.057 |
| miR530/1030 | TCTGCATCTGCACCTGCACCC | pmi-miR530_v1 | 21 nt | 17663 | ppt-miR1030a 5p | 0.004 |
| | TCTGCATCTGCACCTGCACCT | pmi-miR530_v2 | 21 nt | 816 | ppe-miR530 5p | 0.004 |
| miR535 | TGACGACGAGAGAGACGCGC | pmi-miR535_v1 | 21 nt | 19872 | ppe-miR535b 5p | 0.002 |
| miR536 | TCGTGCCAAGCTGTGTCAACC | pmi-miR536_v1 | 22 nt | 26 | ppt-miR536a 3p | 0.002 |
| miR1024 | TCTGGTTGGATTGTAGGCCT | pmi-miR1024_v1 | 20 nt | 21 | ppt-miR1024a 3p | 0.004 |
| miR1083 | TAGCCTGGAACGAAGCACGGA | pmi-miR1083_v1 | 21 nt | 1755 | smo-miR1083 5p | 0.004 |

Reference miRNA refers to a miRBase miRNA that is most similar to a given pmi-miRNA; E-value, complementarity score between pmi-miRNAs and miRBase reference miRNAs as estimated by the BLASTN program.

Ancient land plant miRNA families

MiRBase contains miRNAs from well-characterized plant species for which genome information is available, like the moss *P. patens*, the lycopod *S. moellendorffii* and the angiosperms *Arabidopsis thaliana* and *Oryza sativa* [8, 10], but miRNAs coming from many recent sequencing surveys are not present in the database. To expand the scope of our phylogenetic analysis, we compared the conserved *P. minima* miRNAs to those of *Marsilea quadrifolia* [19], a water fern that diverged from polypod ferns like *P. minima* around 220 MYA [14, 29]. The miRNAs of *M. quadrifolia* and seed plants identified by Chávez-Montes et al. [19] were reanalysed to identify *bona fide* conserved miRNAs (see Materials and Methods). We also compared *P. minima* sequences with the miRNAs of various gymnosperms and basal angiosperms [19, 30–33], as well as conserved miRNAs identified recently in the liverworts *Pellia endiviifolia* [34] and *Marchantia polymorpha* [35, 36]. This allowed us to trace the history of miRNA origin and conservation in land plants, with emphasis on the fern miRNA repertoire.

Particularly noteworthy are miRNA families miR156/319, miR160, miR166, miR171 and miR408, which are found in all land plant groups, including liverworts, which represent the earliest-branching group in land plant evolution (Figs 1A and Fig 3). Some ancestral miRNAs have a more patchy phylogenetic distribution and are missing in a few basal groups. For instance, miR396 is absent from moss, miR167 and miR530 are missing from liverworts and lycopod, while miR535 and miR390 are specifically missing in the lycopod. In any case, the *P. minima* miRNA repertoire is conserved compared to that of basal land plant groups. Thus, all 12 ancient miRNAs present in liverworts (*P. endiviifolia* and/or *M. polymorpha*) and all 15 ancient miRNAs of the moss, are present in *P. minima*. Similarly, all 10 ancient miRNAs found in the lycopod are also present in *P. minima* (Fig 3).

Some miRNA families are only present in a few basal groups. One is miR1024, a high-confidence miRNA from the moss *P. patens* [10] which is found in *P. minima* and in no other species, not even *M. quadrifolia* (Fig 3). Although this miRNA has only 21 reads in our samples, it displays 100% identity over 20 nt to *P. patens* miR1024, indicating that it is a real miRNA. Thus, it might represent an ancient miRNA that has been independently

lost from several lineages, including seed plants. Another case is that of miR1083, which is found in *P. minima* (but not in *M. quadrifolia*, see Materials and Methods) with over 1,700 reads (Table 1). This miRNA is present in the lycopod, *S. moellendorffii* [10], is abundant in gymnosperms [19, 31, 33], but cannot be identified in angiosperms (Fig 3; see Materials and Methods). Thus, miR1083 seems to be an ancient tracheophyte miRNA that has been specifically lost from angiosperms.

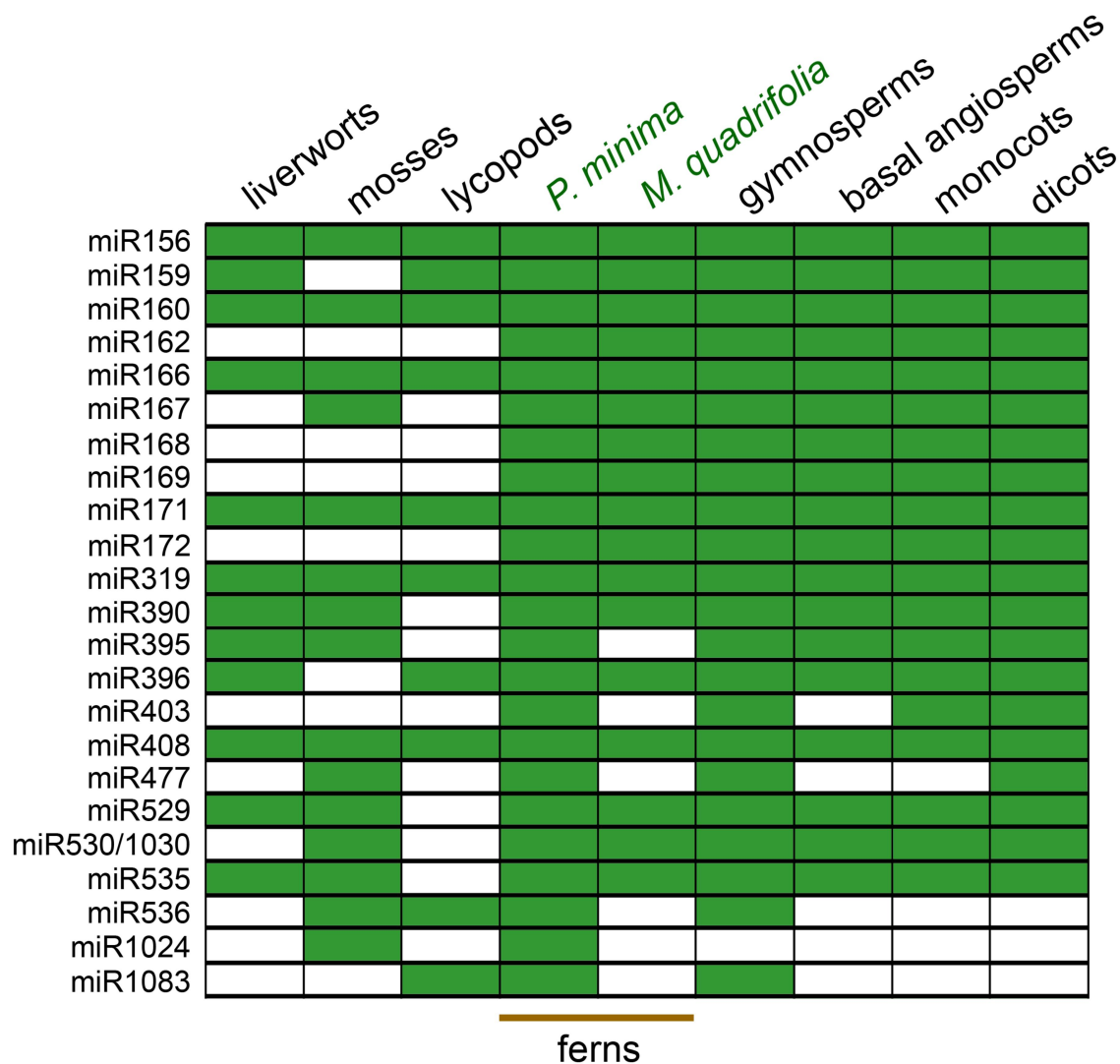


Fig 3. Phylogenetic distribution of conserved fern miRNAs. The grid indicates the presence (green) or absence (white) of miRNA families in liverworts (*M. polymorpha* and/or *P. endiviifolia*), mosses (*P. patens*), lycopods (*S. moellendorffii*), ferns (*P. minima* and *M. quadrifolia*) as well as gymnosperms, basal angiosperms, monocots and dicots.

Euphyllophyte miRNAs

We found that some miRNAs families are only common to ferns (*P. minima*, *M. quadrifolia* or both) and seed plants (gymnosperms and angiosperms), suggesting that these miRNAs originated in the early evolution of euphyllophytes. These are miR162, miR168, miR169, miR172 and miR403 (Fig 3). Among these euphyllophyte miRNAs, the case of miR403 is noteworthy since this miRNA cannot be detected in *M. quadrifolia* nor in most non-dicot groups [19, 37] (see Materials and Methods). However, miR403 has recently been identified in gymnosperms like the Chinese fir, *Cunninghamia lanceolata* [30] and the spruce, *Picea abies* [33], suggesting a complex evolutionary history for this miRNA family.

Conserved fern miRNA targets

Extensive experimental evidence from several angiosperms and the moss *P. patens*, have unveiled that the mRNAs targeted by ancestral miRNAs are often also deeply conserved [7, 10, 11, 38]. Recently, degradome data has shown that the targets of liverwort ancestral miRNAs are also partially conserved with that of other lineages [34-36]. As for ferns, the only available information consists of putative mRNA targets for *C. thalictroides* miR160, miR171 and miR172, which were also found to be conserved with those of other plant lineages [18].

To confirm and expand the evidence for conservation of ancestral miRNA targets in ferns, we computationally predicted the mRNA targets of conserved *P. minima* miRNAs. Since no *P. minima* transcriptome is available, we made use of the extensive transcriptome data obtained for the fern *Lygodium japonicum* [17]. *L. japonicum* belongs to the Schizaeales order, which places it as a distant cousin to polypod ferns like *P. minima* [14, 29]. We compared the conserved *P. minima* miRNA repertoire to the *L. japonicum* transcriptome using the psRNATarget platform [39] and searched for putative targets that had previously been found in other land plants.

The analysis (Table 2) indicates that 11 out of the 23 *P. minima* miRNA families have the potential to target *L. japonicum* RNAs that are also targeted in other species. Thus, pmi-miR156, as well as pmi-miR529, both target a mRNA that encodes a Squamosa

Promoter Binding (SPB) transcription factor; pmi-miR159 and pmi-miR319 both target mRNAs that encode transcription factors of the MYB family; pmi-miR160 targets mRNAs encoding transcription factors of the Auxin Response Factor (ARF) family; pmi-miR166 targets a mRNA encoding a transcription factor of the class III homeodomain-leucine zipper (Class III HD-Zip) family; pmi-miR171 targets a GRAS-domain transcription factor mRNA; miR172 targets the mRNA of a *Apetala-2* transcription factor (Table 2). Targets homologous to these have been observed in experimental work done in angiosperms, moss, or both [10, 11, 38]. The *P. minima* miR160, miR171 and miR172 predicted targets confirm those of Axtell and Bartel [18] for *C. thalictroides*. Recently, a thorough survey found that ferns possess three clades of Class III HD-Zip genes [40], and we found that pmi-miR166 can potentially target all Class III HD-Zip paralogues of the fern *Psilotum nudum* identified in that study (S1 Fig).

Table 2. Putative conserved fern targets of *P. minima* miRNAs.

| miRNA family | <i>Lygodium</i> transcript | E | Transcript function | Target conservation |
|--------------|----------------------------|-----|-----------------------------------------------------|---------------------------------------------------------------------------------------|
| pmi-miR156 | Locus_14997 | 1 | SBP transcription factor | liverworts, moss, gymnosperms, angiosperms |
| pmi-miR159 | Locus_7173 | 1.5 | MYB transcription factor | gymnosperms, angiosperms |
| | Locus_3676 | 2 | RWP-RK domain transcription factor | same target as miR319 in liverworts |
| pmi-miR160 | Locus_6056 | 0 | Auxin response factor (ARF)-B3 DNA binding domain | liverworts, moss, <i>Ceratopteris</i> , gymnosperms, angiosperms |
| | Locus_3364 | 1 | Auxin response factor (ARF)-B3 DNA binding domain | liverworts, moss, <i>Ceratopteris</i> , gymnosperms, angiosperms |
| pmi-miR166 | Locus_1230 | 3.0 | Class III homeodomain-leucine zipper protein C3HDZ2 | liverworts, moss, gymnosperms, angiosperms |
| pmi-miR168 | Locus_158 | 4.0 | Argonaute 1 (AGO1) | gymnosperms, angiosperms |
| pmi-miR171 | Locus_11016 | 1.5 | GRAS domain transcription factor (SCL6) | liverwort (<i>Marchantia</i>), moss, <i>Ceratopteris</i> , gymnosperms, angiosperms |
| pmi-miR172 | Locus_9880 | 0.5 | <i>Apetala2</i> (AP2) | <i>Ceratopteris</i> , gymnosperms, angiosperms |
| | Locus_1423 | 1 | <i>Apetala2</i> (AP2) | <i>Ceratopteris</i> , gymnosperms, angiosperms |
| pmi-miR319 | Locus_7173 | 2.0 | MYB transcription factor | liverwort (<i>Marchantia</i>), moss, gymnosperms, angiosperms |
| | Locus_831 | 3.0 | MYB transcription factor | liverwort (<i>Marchantia</i>), moss, gymnosperms, angiosperms |
| | Locus_1866 | 3.0 | MYB transcription factor | liverwort (<i>Marchantia</i>), moss, gymnosperms, angiosperms |
| | Locus_3676 | 2.0 | RWP-RK domain transcription factor | liverworts (<i>Marchantia</i> , <i>Pellia</i>) |
| pmi-miR390 | Locus_39179 | 0.5 | <i>TAS3</i> tasi-RNA targeting ARF3/4 | liverwort (<i>Marchantia</i>), moss, gymnosperms, angiosperms |
| | Locus_20755 | 0.5 | <i>TAS3</i> tasi-RNA targeting ARF3/4 | liverwort (<i>Marchantia</i>), moss, gymnosperms, angiosperms |
| pmi-miR408 | Locus_5439 | 3.0 | Polyphenol oxydase (PPO) | liverwort (<i>Pellia</i>) |
| pmi-miR529 | Locus_14997 | 1 | SBP domain transcription factor | liverwort (<i>Marchantia</i>), gymnosperms, angiosperms |

E-complementarity score between miRNA and target RNA as estimated by the psRNATarget program

The miR390 family does not control protein-coding mRNAs, but rather targets a group of non-coding transcripts called *TAS3* RNAs (trans-acting siRNAs). In the moss *P. patens* and in angiosperms, two miR390 molecules hybridize to different segments of a *TAS3* RNA, which is then processed by a complex mechanism to produce one or more 21-nt ta-siRNA molecules that are incorporated into a RISC complex that targets the mRNAs of *ARF3* and *ARF4* transcription factors [41]. In addition, *P. patens TAS3* can also produce ta-siRNAs that target the transcription factor *AP2* [10], and there is evidence that the ta-

siRNAs generated from the *TAS3* of the liverwort *M. polymorpha* can target *AP2* as well [36, 42]. As for the *P. minima* miR390 family, we found that they can potentially target two *L. japonicum* non-coding transcripts which exhibit characteristics of *TAS3* RNAs (Table 2). These *L. japonicum* *TAS3* transcripts carry two sites predicted to hybridize to pmi-miR390; in addition, between the miR390 sites one potential ta-siRNA, derived from the (+) strand of the RNA, can be produced that could target *L. japonicum* *ARF3/4* mRNAs (S2 Fig), indicating that the control of *ARF* genes by miR390 is conserved in ferns.

Of particular interest are some predicted *P. minima* miRNA targets that are conserved with liverwort miRNAs but not other plants. Specifically, apart from transcription factors of the MYB family, pmi-miR159 and pmi-miR319 potentially target a mRNA that encode a RWP-RK domain-containing protein (Table 2). Degradome studies in the liverworts *P. endiviifolia* and *M. polymorpha* identified mRNAs encoding RWP-RK proteins as bona fide targets of miR159 and miR319 homologues of these species [34, 35]. Interestingly, the pairing between pmi-miR159/319 and its potential *L. japonicum* RWP-RK target occurs in the 3'-UTR of the mRNA, the same region that is targeted in the homologous liverwort *M. polymorpha* mRNA by miR319 (S3 Fig). Since mRNAs encoding RWP-RK proteins have not been identified as targets of miR159/319 in moss, gymnosperms or angiosperms [10, 11, 30, 31, 33, 38, 43], it is possible that the miR159/319-mediated regulation of RWP-RK proteins is an ancient land plant feature that has been lost from other lineages. Another target possibly conserved between ferns and liverworts is the mRNA encoding the enzyme polyphenol oxydase (PPO), which was identified as a potential target of pmi-miR408 (Table 2). In the liverwort *P. endiviifolia*, *PPO* was also identified as an experimental target of miR408 by degradome analysis [34]. Significantly, miR408 targets the 3'-UTR of *PPO* mRNAs in both ferns and liverworts (S4 Fig) [34].

In seed plants, some miRNAs target genes involved in the biogenesis and function of miRNA themselves. Thus, in angiosperms, miR162 targets Dicer-like 1 (DCL1), miR168 targets AGO1 and miR403 targets AGO2 [11]. In our analyses, *P. minima* miR162 and miR403 are not predicted to target genes similar to DCL1 or AGO2, but pmi-miR168 is predicted to target *L. japonicum* AGO1 (Table2). The sequence predicted to be targeted by pmi-miR168 is located within the 5' half of the *AGO1* coding region, outside the segments that encode the conserved domains of the protein, and corresponds to the

homologous *AGO1* region targeted by miR168 in *A. thaliana* (S5 Fig) [44]. Thus, the negative feedback loop involving miR168 and *AGO1* may be an ancestral euphyllophyte feature.

Discussion

In this work, we identify the conserved miRNAs of a polypod fern, *P. minima*, by comparing its miRNA repertoire to those of other land plant groups. Figure 4 shows a revised scenario for the emergence of different miRNA families during early land plant evolution. Concerning the fern miRNAome, our analyses allowed us to reach a series of conclusions. First, we found that 16 ancient miRNA families that originated in the beginning of land plant evolution, as evidenced by their presence in liverworts and/or mosses, are all present in *P. minima*, showing that the fern miRNA repertoire is very conservative in this regard (Fig 4). This is not a trivial observation, as illustrated by the lycopod *S. moellendorffii*, which lacks half of the ancestral miRNAs that are present in bryophytes (Fig 3). Second, by comparing *P. minima* miRNAs and those of the aquatic fern, *M. quadrifolia*, with the rest of land plants, we identified five miRNAs that seem to have originated at the beginning of euphyllophyte evolution, before the split of the lineages of ferns and seed plants (gymnosperms and angiosperms; Fig 4). The miRNAome of *P. minima* seems to be more conservative than that of *M. quadrifolia*, since six ancestral miRNAs found in the former could be identified in the latter (Fig. 3). Third, by searching putative targets of *P. minima* miRNAs using the transcriptome of another fern, *L. japonicum*, we provide evidence that the targets of 11 fern miRNA families are deeply conserved in land plant evolution.

Molecular studies on basal land plant groups can give crucial clues on the evolution of plant form [3]. As gene expression regulators, miRNAs are essential for proper plant development and may have played important roles in the early evolutionary diversification of land plants, making comparative functional studies of ancestral miRNAs in different plant groups very interesting in this regard [7]. Some ancestral miRNAs present in *P. minima* exhibit an intriguing phylogenetic distribution and are interesting candidates for functional studies. One of these is miR1083, a tracheophyte-specific miRNA that is present

in lycopods and is abundantly expressed in *P. minima* and gymnosperms, but is apparently absent from angiosperms (Figs 3 and 4) [19]. The other is miR1024, which is present in moss and *P. minima*, but has not been detected in any other species investigated so far (Fig 3) [10]. Neither the expression pattern nor the targets of miR1083 and miR1024 have been studied in any plant, but their antiquity and conservation make them interesting subjects for future study.

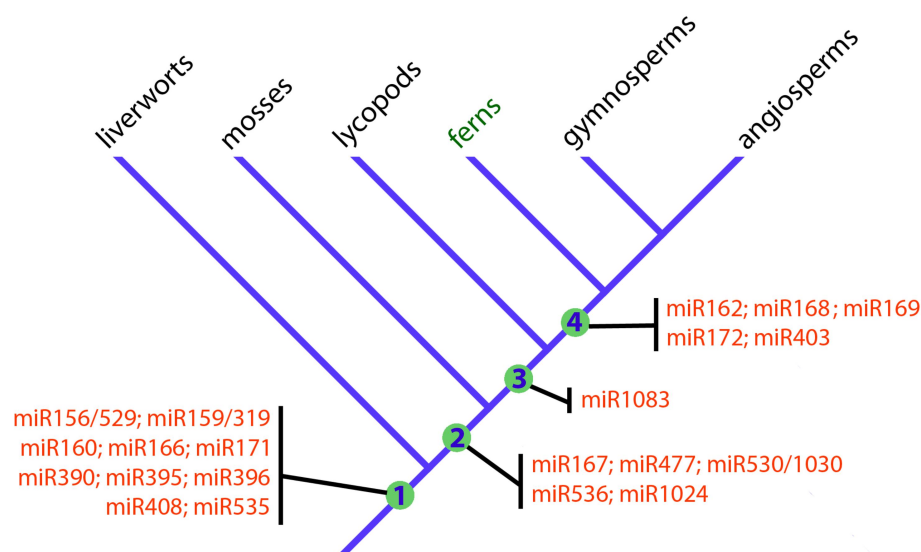


Fig. 4. Emergence of miRNAs during early land plant evolution. The schematic cladogram shows the origin of ancient miRNA families during early land plant evolution. Only miRNA families that originated in early land plant evolution are shown, namely land plants (Embryophyta, node 1), mosses and vascular plants (node 2), vascular plants (Tracheophyta, node 3) and ferns and seed plants (Euphyllophyta, node 4). Note that five miRNA families seem to be restricted to euphyllophytes and that only miR1083 seems to have originated in the branch leading to vascular plants.

Our bioinformatic search for potential targets of fern miRNAs indicate that 11 of the ancestral pmi-miRNAs might target genes that are homologous to those of other land plants (Table 2). Although our evidence is based on an *in silico* approach, the conservation of potential miRNA/target pairs in species that are separated by hundreds of millions of

years of independent evolution is strong evidence that these fern transcripts are real targets. Among these, we found that pmi-miR390 could target *TAS3* RNAs with the potential to generate tasi-RNAs that, in their turn, target genes encoding ARF3/4 transcription factors from *L. japonicum* (S4 Fig). This represents the first full description of *TAS3* transcripts in ferns and an indication that miR390 function is conserved in this group, as suggested by a previous PCR-based survey [42].

Another interesting fern conserved target was found for pmi-miR166, which can target transcription factors of the Class III HD-Zip (C3HDZ) family. In angiosperms, C3HDZ genes are expressed in the adaxial side of developing leaves and are necessary for the proper dorsoventral patterning of these structures [45]. In the abaxial side, on the other hand, C3HDZ expression is repressed by miR166, which is also necessary for proper leaf patterning in angiosperms [46, 47]. Even though it is often considered that the leaves of ferns and seed plants evolved independently [3, 48], Vasco et al. [40] have recently observed that C3HDZ genes are specifically expressed in the adaxial side of developing fern leaves, just like in angiosperms, giving credence to the hypothesis that euphyllophyte megaphylls are homologous structures. Considering our observation that pmi-miR166 could target fern C3HDZ genes, future studies on the expression and function of fern miR166 may help establish the extent of molecular homology in the gene regulatory network that controls leaf development in euphyllophytes.

A recurrent feature in the evolution of plant gene regulation is that certain mRNAs encoding proteins involved in miRNA processing and function are themselves controlled by miRNAs. In angiosperms, *AGO1*, *DCL1* and *AGO2* are targets of miR168, miR162 and miR403, respectively [11], and *AGO1* is also a miR168 target in the Chinese fir, a gymnosperm [30]. Interestingly, liverworts and mosses also target *AGO1* and *DCL1* with a different set of miRNAs. Thus, miR902 and miR1047 target *AGO1* and *DCL1* in the moss *P. patens* [10, 49], while miR11707 targets *AGO1* in the liverwort, *M. polymorpha* [35]. As for ferns, we found that pmi-miR168 has the potential to target *L. japonicum AGO1* in a site that, given its position in the mRNA, seems to be homologous to the *AGO1* site targeted by miR168 in angiosperms. Since miR168 originated in the lineage leading to ferns and seed plants (Fig 4), it can be hypothesized that the negative feedback loop involving *AGO1* and miR168 has been present throughout the evolutionary history of euphyllophytes. The confirmation of the miR168/*AGO1* regulatory relationship, as well as

the study of miR162 and miR403 targets, constitutes an interesting avenue of research in fern regulatory RNA research.

Target prediction of *P. minima* miRNAs yielded two candidate targets that seem to be conserved only between ferns and liverworts. One is the gene encoding polyphenol oxydase (*PPO*), which originated in the early stages of the colonisation of land by plants and has roles in pathogen defence [50]. In both *L. japonicum* and the liverwort, *P. endiviifolia*, *PPO* is targeted at the 3'-UTR by miR408. Another common potential target is *RKD*, a gene encoding a transcription factor of the RWP-RK family. In the liverworts, *P. endiviifolia* and *M. polymorpha*, *RKD* is targeted at the 3'-UTR by miR319 as revealed by degradome analyses [34, 35]. Similarly, we observed that a *L. japonicum* homologue of *RKD* is predicted to be targeted at the 3'-UTR by pmi-miR319 and pmi-miR159. Recently, it has been found that *RKD* is necessary for proper germ cell development in liverworts, as *RKD*-deficient *M. polymorpha* plants exhibit egg cells with aberrant cellular differentiation and proliferation properties, along with defects in gemma cup formation, which are vegetative reproductive organs [51, 52]. Work in *A. thaliana* indicates that *RKD* homologues might have related functions in egg development of seed plants [53]. Interestingly, the overexpression of miR319 in *M. polymorpha* causes defects in thalli growth and gemma cup formation in gametophytes [36]. Thus, given the important functions of *RKD*, its potential regulation by miR319 in liverworts and ferns might represent an ancient gene regulatory interaction that plays a role in gametophyte growth and germ cell differentiation in basal land plants.

Despite their ecological importance and their key phylogenetic position as sister group to seed plants, the genetic and molecular mechanisms underlying fern development and physiology are poorly known. We hope that our work will add to the growing genomic resources available to researchers dedicated to the study of this group of land plants.

Materials and Methods

Plant material

Specimens of *Pleopeltis minima* (Bory) J. Prado & R.Y. Hirai (= *Polypodium squalidum* Vell.; = *Pleopeltis squalida* (Vell.) de la Sota.) [20] were collected in Departamento General Alvear, Corrientes Province, Argentina, near the Uruguay river, in mid-2011. Annual precipitation in the area reaches 17,000 mm approximately. No special permits were needed to collect the specimens. The ferns were found growing on branches of large trees, and a couple of specimens with the branches attached were maintained in a greenhouse at 23°C under a 16 h/8 h light/dark cycle.

RNA preparation and high-throughput sequencing

Two independent samples of *P. minima* fertile fronds from two different specimens each were subjected to a protocol for total RNA extraction enriched for small RNAs using the *mirVana*TM miRNA Isolation Kit (Thermo Fisher Scientific). The quantity and quality of the total RNAs were analysed with a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific) and an Agilent 2100 Bioanalyzer (for concentration, 28S/18S and RIN detection; Agilent Technologies). Small RNA libraries were generated from both *P. minima* independent RNA samples using the Illumina TruseqTM SmallRNA Preparation kit (Illumina, San Diego, USA). Total sRNAs were ligated to 3p and 5p adapters (ADTs), and the corresponding cDNA was obtained by reverse-transcription PCR. The purified cDNA library was used for cluster generation on Illumina's Cluster Station and then sequenced on Illumina GAIIx (LC Sciences, Houston, USA). Raw sequencing reads were obtained using Illumina's Sequencing Control Studio software version 2.8 (SCS v2.8) following real-time sequencing image analysis and base-calling by Illumina's Real-Time Analysis version 1.8.70 (RTA v1.8.70).

Identification of conserved miRNAs

Initial analysis of reads were done with the ACGT101-miR bioinformatics program (LC Sciences, Houston, USA). Raw reads were filtered to remove adaptor sequences and reads which were too short (<15 nt) or too long (>40 nt). Next, sequences corresponding to mRNA fragments, rRNA, tRNA, snoRNA and snRNA mapping to Rfam

(<http://rfam.janelia.org>) were removed, as well as reads mapping to RepeatMasker (<http://www.repeatmasker.org/>). The resulting non-redundant sRNA reads were mapped to plant precursor and mature miRNA sequences stored in miRBase 21.0 (<http://www.mirbase.org/>; Kozomara and Griffiths-Jones, 2014) using BLASTN software to identify conserved miRNAs. Length variation (1- 3 nt) at both 3' and 5' ends and up to two mismatches inside of the sequence were allowed in the alignment. The identified *P. minima* miRNAs were compared to a recent re-evaluation of miRBase entries [8] to discard low-confidence miRNAs from our sample.

To better characterize the phylogenetic distribution of plant miRNAs, *P. minima* sequences were also checked against a large plant miRNA database generated recently [19] which includes the aquatic fern, *M. quadrifolia*. The miRNAs that were reported for *M. quadrifolia* by Chávez-Montes et al [19] were reanalysed more stringently in the following way: First, miRNA sequences for each putative miRNA family were compared against miRBase (<http://www.mirbase.org/>) and miRNA sequences that exhibited low similarity to miRBase entries (E-value of less than 0.01) were discarded. Next, the abundance of each miRNA sequence was checked against the Comparative Sequencing of Plant miRNA database (<http://smallrna.danforthcenter.org/>) [19] and sequences that were expressed at low levels (less than 1 read per million) were discarded. After this conservative reassessment, several miRNAs that seemed to be present in *M. quadrifolia* and other plants [19] were discarded from our evolutionary analysis. The resulting *bona fide* conserved miRNAs were incorporated into the descriptions of miRNA distribution and evolution shown in Figs 3 and 4.

miRNA target prediction

The *P. minima* conserved miRNA sequences identified (57 in total; Table 1) were used to search against the transcriptome of the fern, *Lygodium japonicum* [17], for potential targets using the psRNATarget platform (<http://plantgrn.noble.org/psRNATarget/>) [39]. Searches were performed against the Oases_K49 database (downloaded from <http://bioinf.mind.meiji.ac.jp/kanikusa/>) that consists of a transcriptome assembled from sequenced RNA samples from prothalli, trophophylls, rhizomes and sporophylls obtained

using Roche 454 GSFLX and Illumina HiSeq sequencers [17]. Search parameters were Expectation (E) of ≤ 3.0 (which measures sequence complementarity) and UPE of ≤ 25.0 (which measures target accessibility) [39]. For some miRNAs, namely pmi-miR168, the E parameter was relaxed (≤ 4.0) to identify further potential targets.

Data availability

The raw sequencing data of *P. minima* small RNA transcriptome was submitted to the NCBI Sequence Read Archive under the accession number SRR4294182.

Funding

This study was supported by the National Agency for the Promotion of Science and Technology (ANPCyT-PICT 08-1481) and the National Council for Scientific and Technological Research (CONICET-PIP 11220110100715/12) of Argentina. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgments

The authors thank Dr. Sara Maldonado (University of Buenos Aires) for critically reading the manuscript.

Author Contributions

Conceived and supervised research: AMZ. Performed the experiments: FB PIP LMT. Analysed the data: FSJS SIN AMZ. Contributed reagents/materials/analysis tools: MR ANM. Funding acquisition: ANM AMZ. Wrote the paper: AMZ.

References

1. Kenrick P, Crane PR. The origin and early evolution of plants on land. *Nature* 1997;389:33-39. doi:10.1038/37918
2. Becker B, Marin B. Streptophyte algae and the origin of embryophytes. *Ann Bot.* 2009;103(7):999-1004. doi: 10.1093/aob/mcp044.PMID: 19273476
3. Pires ND, Dolan L. Morphological evolution in land plants: new designs with old genes. *Philos Trans R Soc Lond B Biol Sci.* 2012; 367(1588):508-18. doi: 10.1098/rstb.2011.0252. PMID: 22232763.
4. Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, Shapiro H, et al. The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science.* 2008 Jan 4;319(5859):64-9. PMID: 18079367
5. Banks JA, Nishiyama T, Hasebe M, Bowman JL, Gribskov M, dePamphilis C, et al. The *Selaginella* genome identifies genetic changes associated with the evolution of vascular plants. *Science.* 2011;332(6032):960-3. doi: 10.1126/science.1203810. PMID: 21551031
6. 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 May 30;497(7451):579-84. doi: 10.1038/nature12211. PMID: 23698360
7. Plavskin Y, Timmermans MC. Small RNA-regulated networks and the evolution of novel structures in plants. *Cold Spring Harb Symp Quant Biol.* 2012; 77:221-33. doi: 10.1101/sqb.2013.77.014878. PMID: 23302808
8. Taylor RS, Tarver JE, Hiscock SJ, Donoghue PC. Evolutionary history of plant microRNAs. *Trends Plant Sci.* 2014; 19(3):175-82. doi: 10.1016/j.tplants.2013.11.008. PMID: 24405820
9. Axtell MJ. Classification and comparison of small RNAs from plants. *Annu Rev Plant Biol.* 2013; 64:137-59. doi: 10.1146/annurev-arplant-050312-120043. PMID: 23330790.
10. Axtell MJ, Snyder JA, Bartel DP. Common functions for diverse small RNAs of land plants. *Plant Cell.* 2007; 19(6):1750-69. PMID: 17601824
11. Jones-Rhoades MW. Conservation and divergence in plant microRNAs. *Plant Mol Biol.* 2012; 80(1):3-16. doi: 10.1007/s11103-011-9829-2. PMID: 21996939

12. Sun G. MicroRNAs and their diverse functions in plants. *Plant Mol Biol*. 2012 Sep;80(1):17-36. doi: 10.1007/s11103-011-9817-6. PMID: 21874378
13. Wu G. Plant microRNAs and development. *J Genet Genomics*. 2013 May 20;40(5):217-30. doi: 10.1016/j.jgg.2013.04.002. PMID: 23706297
14. Pryer KM, Schuettpelz E, Wolf PG, Schneider H, Smith AR, Cranfill R. Phylogeny and evolution of ferns (monilophytes) with a focus on the early leptosporangiate divergences. *Am J Bot*. 2004; 91(10):1582-98. doi: 10.3732/ajb.91.10.1582. PMID: 21652310
15. Sessa EB, Banks JA, Barker MS, Der JP, Duffy AM, Graham SW, et al. Between two fern genomes. *Gigascience*. 2014; 3:15. doi: 10.1186/2047-217X-3-15. PMID: 25324969
16. Wolf PG, Sessa EB, Marchant DB, Li FW, Rothfels CJ, Sigel EM, et al. An Exploration into Fern Genome Space. *Genome Biol Evol*. 2015 Aug 26;7(9):2533-44. doi: 10.1093/gbe/evv163. PMID: 26311176
17. Aya K, Kobayashi M, Tanaka J, Ohyanagi H, Suzuki T, Yano K, et al. De novo transcriptome assembly of a fern, *Lygodium japonicum*, and a web resource database, Ljtrans DB. *Plant Cell Physiol*. 2015; 56(1):e5. doi: 10.1093/pcp/pcu184. PMID: 25480117
18. Axtell MJ, Bartel DP. Antiquity of microRNAs and their targets in land plants. *Plant Cell*. 2005; 17(6):1658-73. PMID: 15849273
19. Chávez Montes RA, de Fátima Rosas-Cárdenas F, De Paoli E, Accerbi M, Rymarquis LA, Mahalingam G, et al. Sample sequencing of vascular plants demonstrates widespread conservation and divergence of microRNAs. *Nat Commun*. 2014; 5:3722. doi: 10.1038/ncomms4722. PMID: 24759728
20. Prado J, Hirai, RY. A new combination in *Pleopeltis* and some nomenclatural notes related to illustrations validating fern names. *Am Fern J*. 2010; 100(4):189-194. doi: 10.1640/0002-8444-100.4.189
21. Mai Morente, P. Flora epífita vascular de Uruguay con énfasis en helechos (Polypodiofitas). Tesis de maestría, Universidad de la República (Uruguay). Facultad de Ciencias- PEDECIBA, 2014. Available: <https://www.colibri.udelar.edu.uy/handle/123456789/4055>

22. Proctor MCF, Tuba Z. Poikilohydry and homoihydry: antithesis or spectrum of possibilities? *New Phytol.* 2002 Dec; 156(3):327–49. doi: 10.1046/j.1469-8137.2002.00526.x
23. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* 2014; 42(Database issue):D68-73. doi: 10.1093/nar/gkt1181. PMID: 24275495
24. Zhang SD, Ling LZ, Zhang QF, Xu JD, Cheng L. Evolutionary Comparison of Two Combinatorial Regulators of SBP-Box Genes, MiR156 and MiR529, in Plants. *PLoS One.* 2015 Apr 24;10(4):e0124621. doi: 10.1371/journal.pone.0124621.
25. Morea EG, da Silva EM, e Silva GF, Valente GT, Barrera Rojas CH, Vincentz M, Nogueira FT. Functional and evolutionary analyses of the miR156 and miR529 families in land plants. *BMC Plant Biol.* 2016 Feb 3;16:40. doi: 10.1186/s12870-016-0716-5. PMID: 26841873
26. Palatnik JF, Wollmann H, Schommer C, Schwab R, Boisbouvier J, Rodriguez R, et al. Sequence and expression differences underlie functional specialization of Arabidopsis microRNAs miR159 and miR319. *Dev Cell.* 2007 Jul;13(1):115-25. doi: 10.1016/j.devcel.2007.04.012. PMID: 17609114
27. Li Y, Li C, Ding G, Jin Y. Evolution of MIR159/319 microRNA genes and their post-transcriptional regulatory link to siRNA pathways. *BMC Evol Biol.* 2011 May 12;11:122. doi: 10.1186/1471-2148-11-122. PMID: 21569383
28. Mi S, Cai T, Hu Y, Chen Y, Hodges E, Ni F, *et al.* Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5' terminal nucleotide. *Cell.* 2008 Apr 4;133(1):116-27. doi: 10.1016/j.cell.2008.02.034. PMID: 18342361
29. Schneider H, Schuettpelz E, Pryer KM, Cranfill R, Magallón S, Lupia R. Ferns diversified in the shadow of angiosperms. *Nature.* 2004; 428(6982):553-7. doi: 10.1038/nature02361. PMID: 15058303
30. 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 Aug 15;12:146. doi: 10.1186/1471-2229-12-146. PMID: 22894611
31. 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 Dec;146(4):388-403. doi: 10.1111/j.1399-3054.2012.01668.x. PMID: 22708792
32. 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(5):e0127184. doi: 10.1371/journal.pone.0127184. PMID: 25978425
33. Xia R, Xu J, Arikait 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 Nov;32(11):2905-18. doi: 10.1093/molbev/msv164. PMID: 26318183
34. Alaba S, Piszczalka P, Pietrykowska H, Pacak AM, Sierocka I, Nuc PW, et al. The liverwort *Pellia endiviifolia* shares microtranscriptomic traits that are common to green algae and land plants. New Phytol. 2015; 206(1):352-67. doi: 10.1111/nph.13220. PMID: 25530158
35. Lin PC, Lu CW, Shen BN, Lee GZ, Bowman JL, Arteaga-Vazquez MA, et al. Identification of miRNAs and Their Targets in the Liverwort *Marchantia polymorpha* by Integrating RNA-Seq and Degradome Analyses. Plant Cell Physiol. 2016; 57(2):339-58. doi: 10.1093/pcp/pcw020. PMID: 26861787
36. Tsuzuki M, Nishihama R, Ishizaki K, Kurihara Y, Matsui M, Bowman JL, et al. Profiling and Characterization of Small RNAs in the Liverwort, *Marchantia polymorpha*, Belonging to the First Diverged Land Plants. Plant Cell Physiol. 2016; 57(2):359-72. doi: 10.1093/pcp/pcv182. PMID: 26589267
37. Jagtap S, Shivaprasad PV. Diversity, expression and mRNA targeting abilities of Argonaute-targeting miRNAs among selected vascular plants. BMC Genomics. 2014 Dec 2;15:1049. doi: 10.1186/1471-2164-15-1049. PMID: 25443390
38. Addo-Quaye C, Snyder JA, Park YB, Li YF, Sunkar R, Axtell MJ. Sliced microRNA targets and precise loop-first processing of MIR319 hairpins revealed by analysis of the *Physcomitrella patens* degradome. RNA. 2009; 15(12):2112-21. doi: 10.1261/rna.1774909. PMID: 19850910
39. Dai X, Zhao PX. psRNATarget: a plant small RNA target analysis server. Nucleic Acids Res. 2011; 39(Web Server issue):W155-9. doi: 10.1093/nar/gkr319. PMID: 21622958

40. Vasco A, Smalls TL, Graham SW, Cooper ED, Wong GK, Stevenson DW, et al. Challenging the paradigms of leaf evolution: Class III HD-Zips in ferns and lycophytes. *New Phytol.* 2016 Jul 7. doi: 10.1111/nph.14075. PMID: 27385116
41. Axtell MJ1, Jan C, Rajagopalan R, Bartel DP. A two-hit trigger for siRNA biogenesis in plants. *Cell.* 2006 Nov 3;127(3):565-77. doi: 10.1016/j.cell.2006.09.032 PMID: 17081978
42. Krasnikova MS, Goryunov DV, Troitsky AV, Solovyev AG, Ozerova LV, Morozov SY. Peculiar evolutionary history of miR390-guided TAS3-like genes in land plants. *ScientificWorldJournal.* 2013; 2013:924153. PMID: 24302881
43. Addo-Quaye C, Eshoo TW, Bartel DP, Axtell MJ. Endogenous siRNA and miRNA targets identified by sequencing of the *Arabidopsis* degradome. *Curr Biol.* 2008; 18(10):758-62. doi: 10.1016/j.cub.2008.04.042. PMID: 18472421
44. Vazquez F, Gascioli V, Cr  t   P, Vaucheret H. The nuclear dsRNA binding protein HYL1 is required for microRNA accumulation and plant development, but not posttranscriptional transgene silencing. *Curr Biol.* 2004 Feb 17;14(4):346-51. doi: 10.1016/j.cub.2004.01.035 PMID: 14972688
45. Rodriguez RE, Debernardi JM, Palatnik JF. Morphogenesis of simple leaves: regulation of leaf size and shape. *Wiley Interdiscip Rev Dev Biol.* 2014 Jan-Feb;3(1):41-57. doi: 10.1002/wdev.115. PMID: 24902833
46. Juarez MT, Kui JS, Thomas J, Heller BA, Timmermans MC. microRNA-mediated repression of rolled leaf1 specifies maize leaf polarity. *Nature.* 2004 Mar 4;428(6978):84-8. doi: 10.1038/nature02363. PMID: 14999285
47. Kidner CA, Martienssen RA. Spatially restricted microRNA directs leaf polarity through ARGONAUTE1. *Nature.* 2004 Mar 4;428(6978):81-4. doi:10.1038/nature02366 PMID: 14999284
48. Vasco A, Moran RC, Ambrose BA. The evolution, morphology, and development of fern leaves. *Front Plant Sci.* 2013 Sep 4;4:345. doi: 10.3389/fpls.2013.00345. PMID: 24027574
49. Arazi T. MicroRNAs in the moss *Physcomitrella patens*. *Plant Mol Biol.* 2012 Sep;80(1):55-65. doi: 10.1007/s11103-011-9761-5. PMID: 21373961

50. Llorente B, de Souza FS, Soto G, Meyer C, Alonso GD, Flawiá MM, et al. Selective pressure against horizontally acquired prokaryotic genes as a driving force of plastid evolution. *Sci Rep*. 2016 Jan 11;6:19036. doi: 10.1038/srep19036. PMID: 26750147
51. Koi S, Hisanaga T, Sato K, Shimamura M, Yamato KT, Ishizaki K, et al. An Evolutionarily Conserved Plant RKD Factor Controls Germ Cell Differentiation. *Curr Biol*. 2016 Jul 11;26(13):1775-81. doi: 10.1016/j.cub.2016.05.013. PMID: 27345165
52. Rövekamp M, Bowman JL, Grossniklaus U. *Marchantia* MpRKD Regulates the Gametophyte-Sporophyte Transition by Keeping Egg Cells Quiescent in the Absence of Fertilization. *Curr Biol*. 2016 Jul 11;26(13):1782-9. doi: 10.1016/j.cub.2016.05.028. PMID: 27345166
53. Koszegi D, Johnston AJ, Rutten T, Czihal A, Altschmied L, Kumlehn J, et al. Members of the RKD transcription factor family induce an egg cell-like gene expression program. *Plant J*. 2011 Jul;67(2):280-91. doi: 10.1111/j.1365-3113.2011.04592.x. PMID: 21457369

Supporting Information

S1 Fig. Predicted targeting of fern C3HDZ mRNAs by miR166. (A) Sequence of a transcript (Locus_1230) encoding a class III homeodomain-leucine zipper protein (C3HDZ) homologue from the fern *L. japonicum*. The region predicted to be targeted by pmi-miR166 is indicated in yellow. The coding region is highlighted in blue. (B-D) C3HDZ transcripts belonging to three clades (C3HDZ1, C3HDZ2 and C3HDZ3, respectively) from the fern *Psilotum nudum* [40]. The coding regions and the regions predicted to be targeted by pmi-miR166 are indicated in blue and yellow, respectively.

S2 Fig. *Lygodium japonicum* TAS3 RNAs targeted by miR390. (A) TAS3 RNAs (Locus 20755 and Locus 39179) from *L. japonicum* predicted to be targeted by pmi-miR390. The regions predicted to pair with miR390 are indicated in blue, and the trans-acting small-interfering RNAs (tasi-RNAs) derived from TAS3 processing are shown in yellow. The

tasi-RNAs are predicted to target mRNAs encoding ARF3/4 transcription factors. **(B)** TAS3 RNAs were aligned with the CLUSTAL Omega program. Regions predicted to be targeted by pmi-miR390 are indicated in blue, with the base pairing with miR390 indicated. Note that the pairing between pmi-miR390 is more extensive at the 5' site than at the 3' site, indicating that RISC-mediated cleavage of the RNAs probably occurs only at the 5' site. The trans-acting small-interfering RNAs (tasi-RNAs) derived from TAS3 processing are shown in yellow. The tasi-RNAs are predicted to target mRNAs encoding ARF3/4 transcription factors. **(C)** Sequence of a mRNA (isotig 23398) encoding a ARF3/4 transcription factor from *L. japonicum* predicted to be targeted by tasi-RNAs. The coding region is shadowed grey. The segment highlighted in yellow has almost perfect base-pair complementarity with the tasi-RNAs derived from the TAS3 RNAs shown above.

S3 Fig. *Lygodium japonicum* RWP-RK mRNA targeted by *Pleopeltis* miR159 and miR319. Sequence of a mRNA encoding a RWP-RK transcription factor (Locus 3676) from *L. japonicum* predicted to be targeted by pmi-miR408. The targeted region is located at the 3'-UTR of the mRNA (yellow). The coding region is highlighted in blue. This putative fern miR159/319 target is homologous to an experimental miR319 target of the liverworts, *P. endiviifolia* and *M. polymorpha*.

S4 Fig. Predicted targeting of fern *PPO* mRNA by miR408. Sequence of a mRNA encoding Polyphenol oxydase (*PPO*, Locus 5437) from *L. japonicum* predicted to be targeted by pmi-miR408. The targeted region is located at the 3'-UTR of the mRNA (yellow). The coding region is highlighted in blue. This putative fern miR408 target is homologous to an experimental miR408 target of the liverwort, *P. endiviifolia*.

S5 Fig. Predicted targeting of fern AGO1 mRNA by miR168. **(A)** Sequence of a transcript (Isotig21625) encoding an Argonaute 1 (AGO1) homologue from the fern *L. japonicum*. The region predicted to be targeted by pmi-miR168 is indicated in yellow. The coding region is highlighted in blue. **(B)** Comparison between the experimentally described pairing between ath-miR168a and AGO1 of *A. thaliana* and the predicted pairing between pmi-miR168_v1 and the AGO1 mRNA of *L. japonicum*. The AGO1 sequences recognized

by miR168 are highlighted in yellow. Codons are indicated in alternating red and black lettering and the encoded aminoacid residues are shown below the codons.

S1 Table. Statistics of sRNA libraries.

S2 Table. *P. minima* putative 5p and 3p miRNAs duplexes.