The evolutionary-developmental analysis of plant microRNAs

Sophie Jasinski†, Aurélie C. M. Vialette-Guiraud†
and Charles P. Scutt*

Laboratoire de Reproduction et Développement des Plantes, UMR 5667- CNRS/INRA/Université de Lyon,
Ecole Normale Supérieure de Lyon, 46, allée d’Italie 69364, Lyon Cedex 07, France

MicroRNAs (miRNAs) control many important aspects of plant development, suggesting these molecules may also have played key roles in the evolution of developmental processes in plants. However, evolutionary-developmental (evo-devo) studies of miRNAs have been held back by technical difficulties in gene identification. To help solve this problem, we have developed a two-step procedure for the efficient identification of miRNA genes in any plant species. As a test case, we have studied the evolution of the MIR164 family in the angiosperms. We have identified novel MIR164 genes in three species occupying key phylogenetic positions and used these, together with published sequence data, to partially reconstruct the evolution of the MIR164 family since the last common ancestor of the extant flowering plants. We use our evolutionary reconstruction to discuss potential roles for MIR164 genes in the evolution of leaf shape and carpel closure in the angiosperms. The techniques we describe may be applied to any miRNA family and should thus enable plant evo-devo to begin to investigate the contributions miRNAs have made to the evolution of plant development.

Keywords: microRNA; miR164; angiosperm; flower; carpel; leaf dissection

1. INTRODUCTION

Evolutionary-developmental (evo-devo) biology aims to explain how developmental processes have evolved (Raff 2000). Evo-devo analyses frequently begin from an understanding of gene function in model species and a robust species phylogeny in which those models are placed. Orthologous genes may then be studied in species occupying key phylogenetic positions to deduce the molecular changes that were responsible for evolutionary events of interest. To date, plant evo-devo study has mainly focused on genes encoding developmental regulators such as transcription factors and signal transduction components. However, it is now clear that many plant developmental processes are regulated by microRNAs (miRNAs; Mallory & Vaucheret 2006), suggesting these molecules also may have played important roles in the evolution of plant development.

miRNAs are single-stranded RNA molecules of approximately 21 nt that negatively regulate gene expression by hybridizing to complementary target sites in specific messenger RNAs (mRNAs; Filipowicz et al. 2008). In plants, miRNAs typically show very high similarity to the complement of their target sites in mRNAs, and act by directing the cleavage of these

* Author for correspondence (charlie.scutt@ens-lyon.fr).
†The first two authors contributed equally to this work.


One contribution of 16 to a Discussion Meeting Issue ‘Darwin and the evolution of flowers’.

† The first two authors contributed equally to this work.
functions of MIR164 family members in A. thaliana illustrate the need for evo-devo studies of miRNA genes in plants: the processes of gene duplication, gene loss, sub-functionalization and neo-functionalization, which are known to have shaped families of protein-coding genes (Ohno 1970), have clearly also acted on miRNA families, with potentially important consequences for the evolution of plant development.

Most of the plant miRNA genes currently known were identified by the in silico searching of partial or complete genome sequences (Bonnet et al. 2004; Jones-Rhoades & Bartel 2004; Adai et al. 2005; Li & Zhang 2005; Li et al. 2005). To study the evolutionary importance of miRNAs, methods are required for the identification of miRNA genes in plants occupying key phylogenetic positions whose genomes have not necessarily been sequenced. The high-throughput sequencing of small RNAs (Barakat et al. 2007; Buhtz et al. 2008) can be used to identify mature miRNAs in any species. However, the sequences obtained by this method are too short for phylogenetic analysis, and are therefore of limited value to evo-devo studies.

To help resolve this problem, we present here methods for the efficient identification of miRNA genes in any plant species. To achieve this, we have devised a two-step method, which is specifically applicable to plants due to the generally high conservation of miR sites within any given miRNA gene family throughout the plant kingdom. We furthermore show that standard methods of phylogenetic analysis can be applied to miRNA genes, permitting the reconstruction of miRNA family evolution over considerable phylogenetic distances. As a test case, we have analysed the evolution of the MIR164 family in the angiosperms, or flowering plants. To draw conclusions on the structure of the MIR164 family in the last common ancestor of the extant angiosperms, we incorporated in our studies two representatives of the ANA grade species (ANA for Amborellales, Nymphaeales and Austrobaileyales), which represents the three earliest-diverging lineages of extant angiosperm (Stevens 2001). The ANA grade species used in this study were Amborella trichopoda, which is a shrub endemic to the rainforests of New Caledonia and the only known member of Amborellales, and Cabomba aquatica, which is a small aquatic plant native to northern South America (Orgaard 1991) and a convenient representative of Nymphaeales. We also analysed the MIR164 family in Aquilegia vulgaris, which occupies a key phylogenetic position as a member of the most basally diverging eudicot order, Ranunculales (Stevens 2001).

Our phylogenetic analyses have enabled us to conclude that at least two MIR164 genes were present in the last common ancestor of the extant angiosperms. Furthermore, gene expression analyses across a wide taxonomic range indicate these putative ancestral MIR164 genes to have probably conserved characteristics of their expression patterns throughout angiosperm evolution. We use the results of our analyses to propose testable hypotheses on the roles of miR164 in the evolution of leaf shape and carpel closure in the angiosperms. The methods we have developed should be applicable to any miRNA family and should thus enable studies of the roles that these important regulatory molecules have played in plant evolution.

2. RESULTS
(a) A two-step procedure for the identification of miRNA genes in plants
We have developed a two-step procedure for the identification of miRNA genes of known families in any plant species, as summarized in the electronic supplementary material, figure S1. The first step of this procedure consists of the polymerase chain reaction (PCR) suppression–amplification (Broude et al. 2001) of miRNA gene fragments from adaptor-ligated plant genomic DNA using an miRNA-specific PCR primer in combination with an adapter primer that is common to all DNA molecules. Such amplifications typically yield mixtures of PCR products containing a proportion of miRNA gene fragments. In the second step of our procedure, a database is generated from sequences obtained by PCR suppression–amplification, and then searched for the miR* sites present in miRNA gene fragments. To quantify the efficiency of the PCR suppression–amplification step of our procedure, we used this to re-identify known MIR164 genes from A. thaliana and Zea mays. PCR products generated from genomic DNA of these species were ligated into plasmid vectors, cloned in Escherichia coli, and screened by colony hybridization using MIR164 gene-specific radiolabelled probes. All six MIR164 probes used, Ath-MIR164a, b and c from A. thaliana, and Zma-MIR164a, b and d (Griffiths-Jones et al. 2008) from Z. mays, revealed positively hybridizing bacterial colonies at frequencies of between 0.1 and 8 per cent (see the electronic supplementary material, table S1). This test thus established that our PCR procedure had exhaustively amplified the MIR164 genes present in A. thaliana, in which this family had been completely characterized, and also amplified all three of the MIR164 family members searched for in Z. mays. Despite the success of this initial test, it should not however be assumed that the miRNA gene identification method detailed here will prove to be exhaustive in all cases.

Following the above test, we employed our full, two-step procedure to identify novel MIR164 genes from A. trichopoda, C. aquatica and A. vulgaris. The PCR products inserted into approximately 220 recombinant plasmids, generated from each of these species as described above, were sequenced from one extremity. A DNA database was compiled from sequences of each species under study, and then searched using an optimized BLAST protocol. This procedure resulted in the identification of three putative MIR164 genes from each of A. vulgaris (Avc-MIR164a–c) and C. aquatica (Caq-MIR164a–c), and one such gene from A. trichopoda (Att-MIR164a). The putative miRNA genes identified were present in populations of sequenced plasmids at frequencies of between 0.4 and 2 per cent (see the electronic supplementary material, table S1), in broad agreement with our earlier trial runs using A. thaliana and Z. mays. Further rounds of genomic PCR were then used to complete the sequences of the putative MIR164 genes identified.
The predicted secondary structures and minimal free energy indices (MFEI) of the novel sequences identified (figure 1) are in good agreement with their designation as MIR164 genes (Ambros et al. 2003). PCR amplifications are known to be particularly sensitive to mismatches involving one or both of the two 3′-terminal bases of PCR primers. Accordingly, the PCR amplifications described here were performed with two versions of the miR164 primer, one of which lacked the last two bases of the miR164 consensus sequence. The genes Avu-MIR164a, Caq-MIR164a and Caq-MIR164c were obtained only from amplifications using the shortened PCR primer and were subsequently found to differ from the consensus miR164 sequence at their 3′-extremities (figure 1), indicating the usefulness of this approach. The two-step procedure described here should prove widely applicable for the identification of genes of any miRNA family in any plant species, thereby fulfilling the first essential prerequisite for the evo-devo analysis of plant miRNAs.

(b) A phylogenetic analysis of the MIR164 family in the angiosperms

A multiple alignment was made of the novel MIR164 genes identified in the present study, together with publicly available MIR164 sequences (see the electronic supplementary material, figure S2). Well-conserved nucleotide positions, in and around miR and miR* sites, were selected from this alignment for maximum likelihood (ML) phylogenetic analyses, which were subsequently performed using several different evolutionary models. A typical result of these analyses, obtained using the HKY model, is shown in figure 2. One of the best supported features of our phylogenetic analyses is the monophyly of a group of MIR164 genes we designate as the B-clade, after Ath-MIR164b from A. thaliana. B-clade genes are present in all of the major angiosperm groups analysed, including ANA grade lineages (A. trichopoda, C. aquatica), monocots (Oryza sativa, Sorghum bicolor, Triticum aestivum, Z. mays), basal eudicots (A. vulgaris) and core eudicots (A. thaliana, Populus trichocarpa, Vitis vinifera). The internal structure of the B-clade broadly recapitulates the consensus view of angiosperm phylogeny (figure 2b), except that monocot sequences of the B-clade form a sub-clade that diverges basally to sequences from the ANA grade and eudicots. This observation may reflect a biased sampling of monocots within our study, as all the monocot MIR164 sequences currently available are from the highly derived Poaceae

Figure 1. Secondary structures and minimum free energy indices (MFEI) of pre-miR164s predicted from A. vulgaris (Avu-MIR164a–c), C. aquatica (Caq-MIR164a–c) and A. trichopoda (Atr-MIR164a).
evolution strongly suggests the presence of at least two \textit{MIR164} genes in the last common ancestor of the extant flowering plants. Several other nodes of our phylogeny also appear well supported by high bootstrap values. Such nodes frequently link pairs of genes from closely related species (e.g. \textit{Ath-MIR164a} and \textit{Bna-MIR164a} from Brassicaceae, and \textit{Osa-MIR164d} and \textit{Sbi-MIR164b} from Poaceae), suggesting these genes to be orthologues that were generated by recent speciation events. Other high levels of bootstrap support link multiple \textit{MIR164} genes from a single species (e.g. \textit{Ptr-MIR164b} and \textit{Ptr-MIR164e}, from \textit{P. trichocarpa}), suggesting these genes to be paralogues that were generated by recent duplication events.

\textbf{(c) Expression patterns of MIR164 genes in the angiosperms}

We used semi-quantitative reverse transcriptase-PCR (RT-PCR) to analyse the expression patterns of \textit{MIR164} genes in \textit{A. thaliana}, \textit{C. aquatica}, \textit{A. vulgaris} and \textit{A. trichopoda} (figure 3). B-clade genes of the \textit{MIR164} family were expressed at similar levels in the leaves, flowers and roots of all these species (figure 3a–d), except for a somewhat lower level of expression in the roots of \textit{A. trichopoda} (figure 3d). These data suggest the conservation of similar levels of B-clade gene expression in leaves, flowers and roots of the lineages tested since the initial radiation of the angiosperms. However, further studies on dissected plant tissues, performed using more quantitative methods of analyses, will be required to precisely define the degree of conservation of gene expression pattern between the lineages compared.

A pronounced foliar dimorphism is present in \textit{C. aquatica}, which possesses floating leaves with entire margins that help to carry its inflorescences to the water’s surface, in addition to submerged leaves that are very highly dissected, with the effect of reducing drag in water currents. The B-clade gene \textit{Caq-MIR164a} was more highly expressed in submerged than in floating leaves of \textit{C. aquatica} (figure 3c), while \textit{Caq-MIR164b}, which grouped externally to the B-clade in phylogenetic analyses (figure 2a), was expressed specifically in submerged leaves (figure 3a). Given the known role of \textit{miR164} in determining leaf morphology in \textit{A. thaliana} (Nikovics et al. 2006), these data suggest \textit{miR164} may also be involved in the mechanism generating high levels of dissection in the submerged leaves of \textit{C. aquatica}.

Alone among the sequences studied here, \textit{Caq-MIR164c} from \textit{C. aquatica} showed no measurable expression (figure 3c). Demonstrable expression has been proposed as a prerequisite for the identification of novel miRNA genes (Ambros et al. 2003). However, \textit{Caq-MIR164c} may be considered as a \textit{MIR164} gene as it occupies a well-supported position in phylogenetic reconstructions of the \textit{MIR164} family (figure 2a). This gene may have a very specific expression profile, which was not revealed in the analyses presented here, or may alternatively represent a recently generated pseudogene which is no longer expressed.

\textbf{S. Jasinski et al. miRNA eco-devo analysis}

Mica

vious studies of the evolution of miRNA families thus considerably extends the taxonomic range of pre-
of the extant angiosperms. Our phylogenetic analysis for the identification and phylogenetic analysis of
of the roles these genes have played in plant evolution.
We have demonstrated the efficiency of these pro-
eduction of miRNA genes in plants which should facilitate studies
The roles of MIR164 genes in angiosperm evolution
Leaf lobing and dissection are highly variable traits

Figure 4. MIR164 evolution in the angiosperms. B-clade (red) and one or more other (blue) MIR164 genes were present in the last common ancestor of the extant angiosperms.

conclude that one or more duplications in the MIR164 family occurred before the radiation of the angio-

Our expression analyses (figure 3) have indicated B-clade MIR164 genes to be expressed at a relatively high level in roots, whereas the other MIR164 genes were studies were less expressed in roots than in other tissues. The broad similarity of B-clade gene expression between ANA grade and other angiosperms suggest this pattern to be conserved in the lineages studied since the last common ancestor of the extant angiosperms (figure 4). Similarly, the MIR164 genes external to the B-clade, which may belong to one or more MIR164 lineages, appear in general to have conserved higher expression in leaves and flowers than in roots from an early stage in angiosperm evolution (figure 4).

3. DISCUSSION

(a) Novel methods have permitted a partial reconstruction of MIR164 evolution in the angiosperms

We have devised and tested a combination of methods for the identification and phylogenetic analysis of miRNA genes in plants which should facilitate studies of the roles these genes have played in plant evolution. We have demonstrated the efficiency of these procedures by partially reconstructing the evolution of the MIR164 family since the last common ancestor of the extant angiosperms. Our phylogenetic analysis thus considerably extends the taxonomic range of previous studies of the evolution of miRNA families (Mica et al. 2006; Zhang et al. 2006a). According to the results of our analyses, at least two lineages of MIR164 genes were present in the last common ancestor of the extant angiosperms (figure 4). We therefore conclude that one or more duplications in the MIR164 family occurred before the radiation of the angio-

(b) The roles of MIR164 genes in angiosperm evolution

Leaf lobing and dissection are highly variable traits within the angiosperms, which have clearly arisen many times independently (Bharathan et al. 2002). Such polyphyletic origins for a given morphological trait might imply distinct underlying molecular mechan-
isms. However, the involvement of CUC orthologues in leaf lobing and dissection has recently been demonstrated in several distantly related eudi-
cots, suggesting the independent recruitment of a pre-existing genetic mechanism to these processes in distinct plant lineages (Blein et al. 2008; Berger et al. 2009). In A. thaliana, the effect of CUC2 on leaf lobing is known to be modulated by Ath-MIR164a (Nikovics et al. 2006), suggesting not only the parallel
recruitment of CUC genes to affect leaf morphology, but also of their miR164 regulator. In the present study, we have identified a marked difference in the levels of miR164 between the two leaf forms of C. aquatica, which differ considerably in their degree of dissection. We found miR164 to be more highly expressed in the dissected submerged leaves of this species, which may reflect a need in these organs for a greater modulation of NAC gene activity, related to the production of a highly dissected leaf lamina. Hence, the preliminary results obtained here are consistent with a proposed role for miR164 and its NAC family targets in leaf dissection in an ANA grade angiosperm. A conserved role in leaf dissection for this characteristic is certainly of independent origin, may indicate these genes to function as a genetic module that has been recruited to modify leaf morphorm in distinct lineages from an early stage in angiosperm evolution. Such a module might for example be conserved in all angiosperms, where it would participate in basic morphological processes in meristems, etc. and be ‘switched on and off’ in leaves over the course of evolution to effect relatively ephemeral changes to the lobing or dissection of these organs.

The major defining feature of the angiosperms is the carpel, the specialized female reproductive organ that encloses the ovules. In gymnosperms, which form the sister group to the angiosperms, ovules typically occur as naked structures, borne in the axils of open sporophylls. The process of fusion between carpels is known to be controlled in A. thaliana by the relative expression of miR164 and CUC2 (Baker et al. 2005; Nikovics et al. 2006; Sieber et al. 2007). Further studies may now be undertaken to determine whether a similar mechanism, involving the MIR164 genes identified in the present work, might be responsible for the closure of the simple carpels present in A. trichopoda and C. aquatica. Evidence for the involvement of the miR164/CUC expression balance in carpel closure in multiple ANA grade lineages would suggest a potential mechanism for the initial evolution of the closed carpel in early flowering plants.

4. EXPERIMENTAL PROCEDURES

(a) Plant material

Plants and tissues of A. trichopoda and A. vulgaris were field-collected from Col d’Amieu (New Caledonia) and from Corveissiat (Ain, France), respectively. Seed of the Columbia-0 ecotype of A. thaliana accession no. N1092) was obtained from the Nottingham Arabidopsis Stock Center. Plants of the above species were grown in peat-based compost under greenhouse conditions. Plants of C. aquatica were obtained from Anthias S.A. (Les Chères, France) and grown in a freshwater aquarium.

(b) PCR suppression–amplification of miRNA gene fragments

Genomic DNA was extracted from leaf tissues of A. trichopoda, A. vulgaris, C. aquatica and A. thaliana using a Nucleon PhytoPure kit (Amersham), incorporating an RNAase digestion step. Genomic DNA of Z. mays inbred line A188 (Gerdes & Tracy 1993) was kindly provided by Dr Peter Rogowsky (RDP, Lyon). To amplify miRNA gene sequences, PCR-suppression amplifications were performed, based on the method of Broude et al. (2001). Aliquots of genomic DNA (10 µg) from all species studied were digested with DraI or HaeIII. Double-stranded DNA adaptors (Broude et al. 2001) were ligated to the resulting cleaved DNA fragments, and unincorporated adaptors were then removed using NucleoSpin PCR clean-up columns (Macherey-Nagel). PCR amplifications were performed from aliquots (25 ng) of adaptor-ligated DNA, using a 21-nt primer corresponding to the consensus miR164 sequence from A. thaliana (figure 1) and the A-adaptor primer (Broude et al. 2001). Amplifications were also performed using a 19-mer miR164 primer, lacking the two bases at the 3’-extremity of the consensus miR164 sequence. PCR amplifications were performed using 0.02 units µl⁻¹ of EX-TAQ Polymerase (Takara), 0.2 mM of each primer, 0.2 mM dNTPs and reaction buffer (1 ×), as supplied by the manufacturer, and continued for 30 thermal cycles, each consisting of a denaturation step of 94°C for 30 s, a thermal gradient annealing step of 45–65°C for 30 s, and an elongation step of 72°C for 90 s. Aliquots of the PCR products generated were analysed on 1 per cent agarose gels. DNA samples for further study were selected from amplifications performed at the highest annealing temperature that yielded products in each case. Aliquots (1 µl) of the DNA samples selected were ligated into the pCRll vector (Invitrogen) and the recombinant plasmids generated were used to transform E. coli DH5α competent cells and plated on ampicillin-containing media.

To estimate the proportion of MIR164 gene fragments in PCR products amplified from A. thaliana and Z. mays genomic DNA, 520 and 1040 ampicillin-resistant colonies generated, respectively, from these two species were transferred onto Hybond-N (Amersham) membranes for colony hybridizations using radio-labelled, gene-specific probes of Ath-MIR164a, b and c (for A. thaliana genes), or Zma-MIR164a, b and d (for Z. mays genes). To identify novel MIR164 genes from A. trichopoda, C. aquatica and A. vulgaris, approximately 220 antibiotic-resistant bacterial colonies from each species, generated as described above, were sent to a commercial service provider for plasmid purification and DNA sequencing using the M13 forward sequencing primer. PCR products derived from DNA cleaved by DraI and HaeIII, and from amplifications involving full-length and shortened miR164 primers, were selected for sequencing in approximately equal proportions.

(c) Database construction and BLAST searching for miR* sites

DNA sequences of PCR products, generated as described above, were imported into CONTIG EXPRESS (Invitrogen). Cloning vector regions were removed, and the resulting sequences were assembled into
contigs. For contigs containing a miR164 primer at the 5′-end, only the longest contributing sequence was retained for further analysis, whereas for contigs beginning with an adaptor primer, the whole contig was saved, reversed and complemented. The resulting sequences from each plant species were exported as a FASTA file and aligned over their terminal miR164 primer sequences, where present, using MUSCLE (Edgar 2004), which allowed these primer sites to be removed easily from the aligned sequences. The resulting primer-less sequences were then compiled into stand-alone DNA database using FORMATDB. The DNA databases generated in this way were searched by BLAST (Altschul et al. 1997) for internal miR* sites using the mature miR164 consensus sequence (figure 1). BLAST options were used to specify searching of the bottom strand only (S2), with a word-size of 6 (W6), no filtering (FF), giving a score of 4 for an exact match (r4), and a penalty of 5 for each mismatch (q-5). BLAST hits were used to search the available plant nucleotide databases to eliminate any false positives that were homologous to known genes. The remaining sequences, representing putative MIR164 genes, were retained for further study.

(d) PCR amplification across miR sites

To determine the sequences of mature miR sites present in the putative MIR164 genes identified, and of the regions immediately adjacent to these, a second round of PCR suppression–amplification was undertaken. These amplifications were performed on further aliquots of adaptor-ligated genomic DNA by the sequential use of two miRNA gene-specific primers derived from the novel sequences obtained, in combination with the A-adaptor primer (Broude et al. 2001). The resulting PCR products were cloned in plasmid vectors and sequenced, and the novel data obtained were used to complete the sequences of the putative MIR164 genes identified.

(e) RNA secondary structure prediction and stability calculations

Secondary structures of putative pre-miR164 molecules were predicted in MFOLD 3.2 (Zuker 2003) using default parameters. MFEI for the structures generated were calculated as described by Zhang et al. (2006b).

(f) Phylogenetic analyses of miRNA genes

Sequence alignments were performed using M-Coffee (Wallace et al. 2006) using ‘regular’ parameters and homologous sites were chosen for ML phylogenetic analyses on the basis of their clear alignment. The full alignment is shown for clarity in the electronic supplementary material, figure S2, though only the sites chosen for phylogenetic analyses in this figure may be considered as homologous. Phylogenetic analyses were performed using TREEFINDER v. May 2006 (Jobb et al. 2004), initially assuming the HKY (Hasegawa et al. 1985) model of DNA substitution. Congruent results were also obtained assuming the GTR (Rodriguez et al. 1990) and TrNef+f+Γ (Tamura & Nei 1993) models. In each analysis 1000 bootstrap replicates were performed.

(g) Semi-quantitative reverse transcriptase-PCR analyses

Plant tissues were ground under liquid nitrogen, together with 0.1 g of polyvinylpyrrolidone (mw 40 kDa) per gram of tissue, and RNA was extracted from these tissue powders using an RNA-Easy Kit (Qiagen). Samples (0.5 μg) of total RNA were then treated with RNase-free DNase (Ambion), and each divided into two aliquots. These aliquots were processed for first-strand cDNA synthesis, respectively with and without the addition of RevertAid M-MuLV RT (Fermentas), using oligo-dT primers and other reaction components as described by the manufacturer. Appropriate volumes of diluted cDNA samples, and equal volumes of similarly diluted negative control samples (not treated with RT), were used as templates in PCR amplifications using MIR164 gene-specific and GAPDH primers. Full details of the primer sequences and thermal cycle conditions used are given in the electronic supplementary material, table S3.

We thank Patrick Laufs for helpful discussions and for supplying Arabidopsis miR gene-specific probes, Manolo Gouy and Dominique Mouchiroud for advice on phylogenetic analysis, Guy Perrière for advice on database methods, and the technical staff of the RDP Laboratory for their support. S.J. was supported by an ATER post of the ENS-Lyon. A.V. receives a doctoral thesis grant from the French Ministry of Research.

REFERENCES


