

SCIENTIFIC REPORTS



OPEN

In silico identification and characterization of a diverse subset of conserved microRNAs in bioenergy crop *Arundo donax* L.

Wuhe Jike^{1,2}, Gaurav Sablok^{1,3}, Giorgio Bertorelle², Mingai Li¹ & Claudio Varotto¹

MicroRNAs (miRNAs) are small non-coding RNA molecules involved in the post-transcriptional regulation of gene expression in plants. *Arundo donax* L. is a perennial C₃ grass considered one of the most promising bioenergy crops. Despite its relevance, many fundamental aspects of its biology still remain to be elucidated. In the present study we carried out the first *in silico* mining and tissue-specific characterization of microRNAs and their putative targets in *A. donax*. We identified a total of 141 miRNAs belonging to 14 families along with the corresponding primary miRNAs, precursor miRNAs and a total of 462 high-confidence predicted targets and novel target sites were validated by 5'-race. Gene Ontology functional annotation showed that miRNA targets are constituted mainly by transcription factors, but three of the newly validated targets are enzymes involved in novel functions like RNA editing, acyl lipid metabolism and post-Golgi trafficking. Folding variability of pre-miRNA loops and phylogenetic analyses indicate variable selective pressure acting on the different miRNA families. The set of miRNAs identified in this study will pave the road to further miRNA research in *Arundo donax* and contribute towards a better understanding of miRNA-mediated gene regulatory processes in other bioenergy crops.

MicroRNAs (miRNAs) are endogenous small non-coding RNA molecules, containing approximately 22 nucleotides (nt), playing important roles in the regulation of gene expression at the post-transcriptional level¹. In plants, miRNA genes are transcribed by RNA polymerase II into primary miRNAs (pri-miRNA)², long precursors displaying polyadenylation like protein-coding mRNAs. In the nucleus, pri-miRNAs are processed by the Dicer-like 1 (DCL1) enzyme into shorter stem-loop hairpins, called precursor miRNAs (pre-miRNAs). After transportation into the cytoplasm by exportin 5, the pre-miRNAs are further cleaved by DCL1 to produce a duplex formed by the mature miRNA and its star miRNA (miRNA*), the nearly perfect reverse complementary RNA derived from the pre-miRNA stem³. Subsequently, the single strand of mature duplex corresponding to the mature miRNA is assembled with an Argonaute (AGO) RNA binding protein to form the RNA-induced silencing complex (RISC), which facilitate the interaction of mature miRNAs with their target mRNAs^{4,5}. RISC-associated plant miRNAs recognize their target mRNA sequences by their nearly perfect or perfect complementarity, allowing them to identify with extremely high specificity only a small fraction of all transcribed mRNAs. This very high specificity is the key to enable microRNAs to regulate the expression of their targets. For the majority of plant miRNAs, target gene expression regulation is achieved by transcript cleavage, usually occurring between the 10th and 11th nucleotide at the 5' end of the miRNA⁶. However, translational inhibition can be an additional/alternative mechanism used by some microRNAs to downregulate target expression^{7,8}. In addition, the ability of a single miRNA to be potentially involved in the regulation of multiple target genes or of multiple miRNAs⁹ makes microRNAs very flexible regulators in a wide variety of metabolic and biological processes during all major growth and developmental stages of plants¹⁰.

¹Department of Biodiversity and Molecular Ecology, Research and Innovation Centre, Fondazione Edmund Mach, San Michele all'Adige, Trento, Italy. ²Università degli Studi di Ferrara, Dipartimento di Scienze della Vita e Biotecnologie, Ferrara, Italy. ³Department of Biosciences, University of Helsinki, Helsinki, Finland. Correspondence and requests for materials should be addressed to M.L. (email: mingai.li@fmach.it) or C.V. (email: claudio.varotto@fmach.it)

Modern high-throughput sequencing technologies hold great promise to produce large sets of genomic or transcriptomic data in different tissues and at different developmental stages, which provides useful sequence resources to predict and analyze miRNAs in non-model plant species with non-sequenced genomes. In plants, computational approaches have been successfully applied and demonstrated effective to attain a comprehensive prediction of potential miRNAs, such as in *Cassava*, strawberry, *Arabidopsis thaliana* and many others^{11–15}. Besides, the existence of curated online databases like miRBase¹⁶, collecting and organizing in a reference repository all miRNAs predicted by computational approaches, enormously simplifies microRNA identification in novel transcriptomes. Both homology searches based on miRNA conservation among different plant species and the secondary hairpin loop structures of the pre-miRNA sequences along with the high negative minimal folding energy (MFE) are reliable criteria for the computational identification of miRNAs. This is why, even without experimental validation, miRNAs can reliably be distinguished from other types of small RNAs, thus reducing the number of false positive among predicted miRNAs^{17,18}. Especially in plant species, the feature of nearly perfect or perfect complementarity of miRNAs to their target mRNA sequences also allows the reliable computational prediction of miRNA target genes. This, in turn, is very important for *in silico* prediction of miRNA functions, which allowed the genome-wide identification of microRNA genes^{19–21}.

Arundo donax L., also called “giant reed”, is a perennial C₃ fast growing grass²². Genetic studies indicate that *A. donax* originated in Eastern Asia, from where it spread, possibly by human intervention, to the Middle East and the Mediterranean²³. More recently it was introduced in Africa and even Australia. In the large majority of its distribution area *A. donax* is reported to be a sterile species reproducing vegetatively, and it has been suggested that it may be a hybrid with uneven ploidy or possibly a (pseudo-) triploid species^{24,25}. Despite its sterility, the vigorous growth and lack of natural antagonists allowed *A. donax* to become one of the most invasive riparian species in Southern USA (especially California), where it was introduced approximately one century ago to consolidate the sides of rivers²⁶. This robustness makes it a very productive biomass species, as in optimal conditions *A. donax* fields become productive already after the second year and can provide dry biomass yields up to 40 tons per hectare for the next ten years. These yields are higher than those of other perennial rhizomatous grasses, thus constituting one of the most promising species for the production of biomass in the Mediterranean area²⁷. In addition, *A. donax* requires little management input, as it is resistant to most pests and pathogens. It can grow without significant P or N fertilization, and is highly tolerant to heavy metals and saline soil^{28–30}. *A. donax* has also been utilized as a raw material for bioethanol production with diluted oxalic acid pre-treatment, which is important to overcome recalcitrance of lignocellulose for ethanol production³¹. Recently, several transcriptomic studies have started to elucidate the content of expressed genes in *A. donax* (e.g.^{32–34}), providing the opportunity for the in-depth mining of its gene space.

In the present study, we carried out the first computational identification and characterization of miRNAs for the biomass and bioenergy crop *A. donax* using tissue-specific transcriptomic data, as this could provide novel insights into the mechanism of *A. donax* development, metabolism and biology. We also predicted their putative target transcripts, providing through network analysis an in-depth dissection of gene ontologies and functional annotations for both putative miRNAs and target genes. These findings advance our understanding of miRNAs in *A. donax*, and have the potential to be further utilized for controlling secondary metabolism for improving the production of biomass and fermentation efficiency.

Methods

Prediction of potential *A. donax* miRNAs. All previously known 1616 miRNA precursor sequences from 12 monocotyledon species were downloaded from the miRBase database³⁵ (Release 21.0; <http://www.mirbase.org/> (2017)). These precursor miRNAs were used as query sequences for BLASTN searches against the four reference bud, culm, leaf and root transcriptomes of *A. donax*³² using default parameters and an E-value cut-off of 10. Only the best hit for each query sequence was retained and after elimination of redundant hits, these candidate primary miRNA sequences were scanned for hairpin-like secondary structures using the miRNA identification pipeline of the C-mii software³⁶ (Supplementary Fig. 1). For prediction, only the miRNAs of *Oryza sativa* were used as reference, as the annotation of microRNA genes in this species is by far the most complete and reliable. To reduce type I (false positive) errors at the possible expense of a somehow inflated number of false negatives, we applied a stringent filtering of the primary microRNAs (pri-miRNAs) identified by C-mii (Supplementary Fig. 1). Only candidate sequences fitting the following criteria were considered as putative miRNAs in *A. donax*: (1) The length of predicted mature miRNAs should be in the range of 19–25 nucleotides; (2) A maximum of two mismatches compared with known rice mature miRNAs should be allowed; (3) The mature miRNA should be localized in only one arm within the predicted stem–loop structure; (4) No more than five mismatches should be allowed between miRNA sequence and guide miRNA sequence in the stem–loop structure; (5) miRNAs should have high A + U content (30–70%); and (6) minimal folding free energy (MFE) and minimal free energy index (MFEI) value of the secondary structure should be highly negative, with a cut-off value of -0.85 kcal/mol^{10,37,38}. *A. donax* putative microRNAs were renamed according to the closest homologous locus in rice, identified as the best hit in BLASTN searches against all rice pre-miRNAs. Clustering of *A. donax* pri-miRNAs was finally carried out to identify tentative genetic loci, as no reference genome sequence is available for this species. Sequences with less than two mismatches in BLASTN searches over the whole alignment length were considered alleles of the same locus and renamed accordingly.

Position-specific base composition of mature microRNAs. Nucleotide composition and their dominance at particular positions in mature *A. donax* miRNAs and reference *O. sativa* were analyzed using BioEdit³⁹. Base composition frequency were calculated for each position of *A. donax* and *O. sativa* mature miRNAs, the average percentage of A, C, G and U bases was then calculated across all families. The position-specific nucleotide frequency of predicted mature miRNAs was summarized in graphical form.

Structural and phylogenetic reconstruction of different microRNA families. The precursor sequences of the identified *A. donax* miRNAs were further analyzed to investigate stem-loop structure variabilities of pre-miRNAs. MUSCLE was used to align sequences which were subsequently used for phylogenetic analysis in MEGA 7.0 by employing the neighbor joining method with 1000 bootstrap replicates⁴⁰.

Prediction and functional annotation of putative *A. donax* miRNA targets. Putative microRNA targets were identified with two different programs: psRNATarget⁴¹ and TargetFinder⁴². The parameter sets for prediction by the psRNATarget server (<http://plantgrn.noble.org/psRNATarget/home>) were: maximum expectation of the score between small RNAs and their target transcripts: 3.0; complementarity scoring length (hsp-size): 20; maximum allowed unpaired energy (UPE): 25; flanking length for analysis of target accessibility: 17 nt upstream and 13 nt downstream of the target site; central mismatch range leading to translation inhibition: 10-11 nt. Prediction of candidate targets with a stand-alone version of the TargetFinder program was carried out with default parameters. Sequences with a score lower than 4 were regarded as predicted miRNA target genes. The transcripts identified by both programs after removing non-coding transcripts by BLAST similarity search against the non-redundant protein databases from NCBI protein database were considered as putative microRNA targets and used for subsequent analyses. To better understand the function of *A. donax* miRNAs and their regulating targets, Gene Ontology (GO) annotation of the predicted *A. donax* miRNA targets were performed by using the annotation web tool FunctionAnnotator (<http://fa.cgu.edu.tw/index.php>)⁴³. Further functional annotation of the predicted targets was carried out performing BLASTX searches against the *Arabidopsis thaliana* (<https://www.arabidopsis.org/> (2018)) and *Setaria italica* (<http://www.uniprot.org/> (2018)) protein databases using default parameters and an E-value cut-off of $1e^{-5}$. The biological networks formed by the putative miRNAs and their targets were visualized by Cytoscape version 3.5⁴⁴.

Total RNA extraction and cleavage site identification of miRNA targets. To confirm and identify miRNA-directed cleavage sites on target genes, total RNA isolations from root and bud in *A. donax* were performed as formerly described³⁴. 5' RACE System for Rapid Amplification of cDNA Ends, Version 2.0 kit (Invitrogen) was used for 5' race experiments according to manufacturer's instructions. The PCR products amplified using second nested primers were gel purified, cloned into pGEM-T (Promega) and sequenced. All primers used are listed in Supplementary Table 1.

Comparative genomic analyses of conserved miRNA targets in *Arundo donax* and other plants. All predicted targets of the 11 conserved miRNA families from *Oryza sativa*, *Zea mays*, *Arabidopsis thaliana* and *Vitis vinifera* were downloaded from the PNRD database (<http://structuralbiology.cau.edu.cn/PNRD>) (Yi *et al.*, 2015) and used for TBLASTN and BLASTX searches against the putative *Arundo donax* targets with an E-value cut-off of $1e^{-5}$. Hits with a Score value greater than 50 and with sequence coverage to the query greater than 50% (R. Pearson, 2013) were retained as conserved homologs, while the others were considered novel targets.

Results

Identification of putative miRNAs in *A. donax* and their characteristics. Through blast searches of the reference transcriptome of *A. donax* (1,195,562 transcript sequences) and microRNA prediction with the C-mii program we identified a total of 310 miRNA candidates. For reducing the false positives and improving the accuracy of the prediction, we retained for subsequent analyses only the predicted pre-miRNA with highly negative values of *MFEI* (< -0.85 kcal/mol). In this way, we identified a total of 141 high-confidence putative miRNAs belonging to 14 different families (Supplementary Table 2, Supplementary sequences), corresponding to the most common miRNA families in *O. sativa*. *A. donax* putative miRNAs varied from 20 to 22 nucleotides in length, with the majority of them being 21 nt in length (85.82%), followed by 20 nt (7.80%), and 22 nt (6.38%), respectively. The lengths of precursor miRNAs varied from 60 to 193 nt with an average value of 99 nt, in line with what has been found in other plant species (Supplementary sequences). The *Ado-MIR444d-1c_b*, *Ado-MIR444d-2c_b*, *Ado-MIR444d-2c_c*, *Ado-MIR444d-2c_l*, *Ado-MIR444d-3c_c*, *Ado-MIR444d-4c_l* and *Ado-MIR444d-5c_r* exhibited the shortest precursor length of 60 nt, whereas *Ado-MIR169n-2_r* showed the longest precursor length of 193 nt (Supplementary Fig. 2 and Supplementary Table 2).

Among the 14 miRNA families, 10 (MIR166, MIR396, MIR529, MIR827, MIR160, MIR319, MIR1430, MIR167, MIR171 and MIR172) contained one to nine members, while the remaining four families (MIR156, MIR169, MIR393 and MIR444) were found to have more than ten members. The MIR444 family was the largest family with 55 members (Supplementary Table 2 and Fig. 1). We used the 94 miRNA loci of *O. sativa* corresponding to the 14 families from miRBase³⁵ as reference to reliably identify tentative miRNA loci in *A. donax*. Based on the number of paralogs present in rice, a total of 69 loci were identified in *A. donax*. The MIR169 family had the highest number of loci both in *O. sativa* and *A. donax*. 18.75% of *A. donax* MIR169 loci corresponded to single loci in *O. sativa*. The highest number of *A. donax* loci per rice gene were seven, in line with the polyploidy of the giant reed (Table 1). 26.24% of the miRNAs generated from unique loci and primary transcripts, e.g., locus_17 and locus_18 (Supplementary Table 2). The miRNA identified from these loci did not show marked tissue-specific preferences, with 29 loci expressed in roots, followed by 20, 17 and 15 in culms, buds and leaves, respectively. However, 13.04% of loci generated multiple primary transcripts and miRNAs. For instance, there were three transcripts corresponding to locus_63. The miRNAs identified from this locus belong to the MIR444 family, and these miRNAs were expressed in buds, culms and leaves; another example was the two transcripts from locus_69, which belonged to MIR827 family and they were expressed in buds and roots (Supplementary Table 2).

By analyzing more in-depth the tissue-specific co-expression pattern, we found that most of the identified miRNA families were preferentially expressed in root (31.9% of the mature miRNAs; Fig. 2), followed by

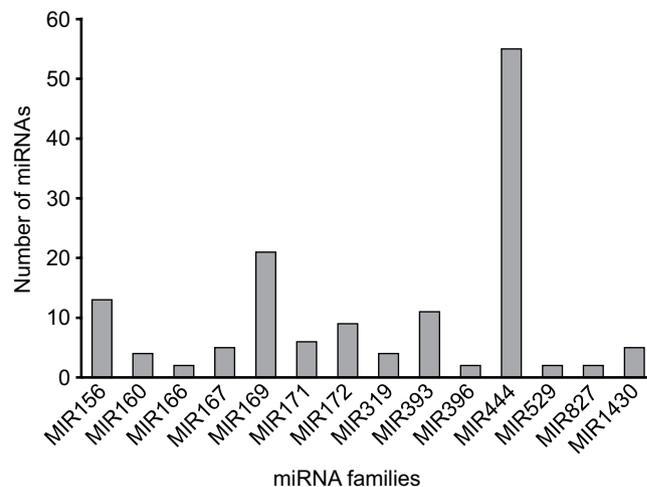


Figure 1. Number of miRNA member (s) predicted for each of 14 families identified in *A. donax*.

miRNA family	loci in <i>Oryza sativa</i>	loci in <i>Arundo donax</i>	loci in common [#]
MIR156	12	8	a(1), b(1), c(1), g(3), j(1), k(1)
MIR160	6	4	b(3), c(1)
MIR166	13	2	a(2)
MIR167	10	5	d(2), g(3)
MIR169	18	16	a(1), c(3), i(1), n(3), p(1), q(7)
MIR171	9	5	a(1), c(2), f(1), i(1)
MIR172	4	3	b(1), d(2)
MIR319	2	2	a(2)
MIR393	2	4	a(1), b(3)
MIR396	8	2	a(1), b(1)
MIR444	6	12	a(2), c(4), d(5), e(1)
MIR529	2	1	a(1)
MIR827	1	1	*(1)
MIR1430	1	4	*(4)

Table 1. Rice homologs of *A. donax* microRNA loci. [#]Letters correspond to the names of single loci in *O. sativa*, numbers in brackets correspond to the number of inferred loci in *A. donax*. *Single locus in *O. sativa*.

culm (26.2%), bud (22.7%) and leaf (19.1%). Among the conserved miRNAs, about one third of the families were expressed in all four tissues studied, namely MIR444 (the largest family distributed on a per-tissue basis), MIR169, MIR167, MIR393 and MIR172. Only a minority of the sampled families were expressed only in one tissue, namely MIR166, which showed specific expression in the root, and MIR396 and MIR529, which were specifically expressed only in the culm (Supplementary Table 3). Overall, a relatively moderate differential expression was observed for all the predicted miRNA in the four tissues. MFE is an important parameter for determining the reliability of secondary structures of pre-miRNA, as the stability of the stem-loop structures of the precursor miRNAs is more stable when MFE has highly negative values. In the present study, the range of MFE (-kcal/mol) calculated was -26.4 to -81.8 (kcal/mol) with an average value of -48.67 (kcal/mol). MFEI is the minimal folding energy index, which can be used to distinguish pre-miRNA from other coding or non-coding RNA and RNA fragments. MFEI values ranged from -0.85 (kcal/mol); the maximal cut-off used for prediction) to -1.402 (kcal/mol) with an average of -1.03 (kcal/mol). These values were significantly lower than other reported small RNAs such as tRNAs (-0.64 kcal/mol), rRNAs (-0.59 kcal/mol) and mRNAs (0.62–0.66 kcal/mol), indicating that the identified *A. donax* miRNAs were putative miRNAs with high confidence^{18,45,46} (Supplementary Table 2).

Analysis of position-specific nucleotide preference in *A. donax* mature miRNAs. The overall percentage of each base of mature *A. donax* miRNAs was found to be 28.46% for uracil, 24.54% for cytosine, 25.62% for guanine and 21.39% for adenine (Supplementary Table 4). These values were in line with base compositions in *Oryza sativa* mature miRNAs. In general, a slightly lower GC content was apparent in *A. donax* (50.15%) than in *O. sativa* (50.69%) (Fig. 3a; Supplementary Table 4).

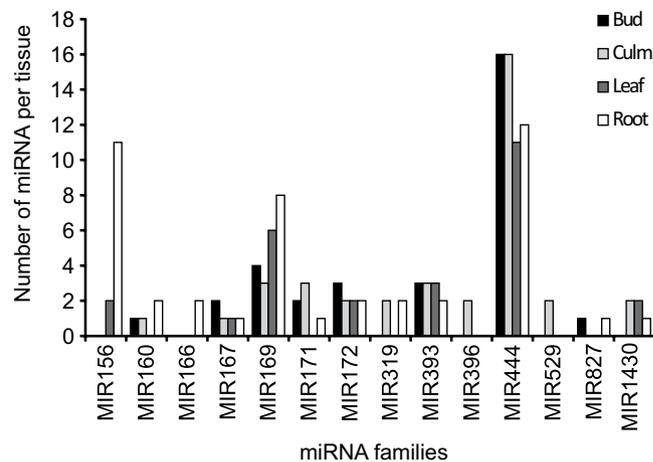


Figure 2. Number of mature miRNAs present in different tissues (bud, culm, leaf and root) of *A. donax*.

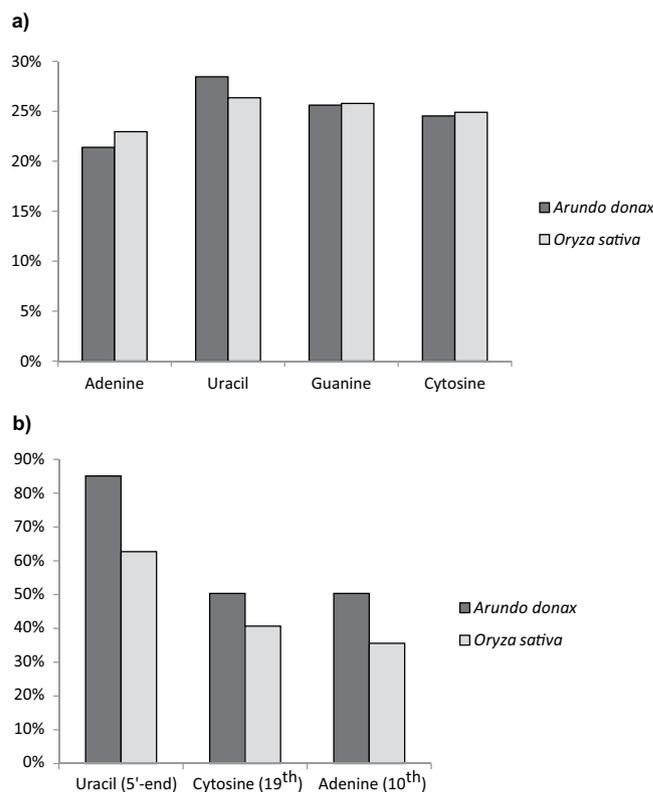


Figure 3. Composition of nucleotides in mature miRNAs from *A. donax* and *O. sativa*. (a) Overall nucleotide compositions (%) compared among mature miRNAs. (b) Selected position-specific nucleotide compositions of mature microRNAs in *A. donax* and *O. sativa*.

In the 5'-end of *A. donax* miRNA, uracil was found in 85.11% of the sequences, while in *O. sativa* it was present in 62.71% of the cases. Also other positions showed different base preferences as compared to rice. Cytosine was found to be a dominating base at position 19 (50.35%) in *A. donax*, while in rice it was present in only 40.68% of the cases. Adenine (50.35%) was abundant at the 10th nt position of *A. donax* mature miRNAs, but less abundant in the same position in rice (35.59%; Fig. 3b; Supplementary Table 4). We notice, however, that the differences observed here only refer to the 14 microRNA families identified in *A. donax* with high conservation to rice and highly stringent selection parameters, which most likely do not provide a comprehensive picture of all microRNAs present in *A. donax*.

Variability of stem-loop structures in *A. donax* pre-miRNAs. To understand folding variability of stem-loop structures and phylogenetic relationship among *A. donax* pre-miRNAs, the stem-loop structures of

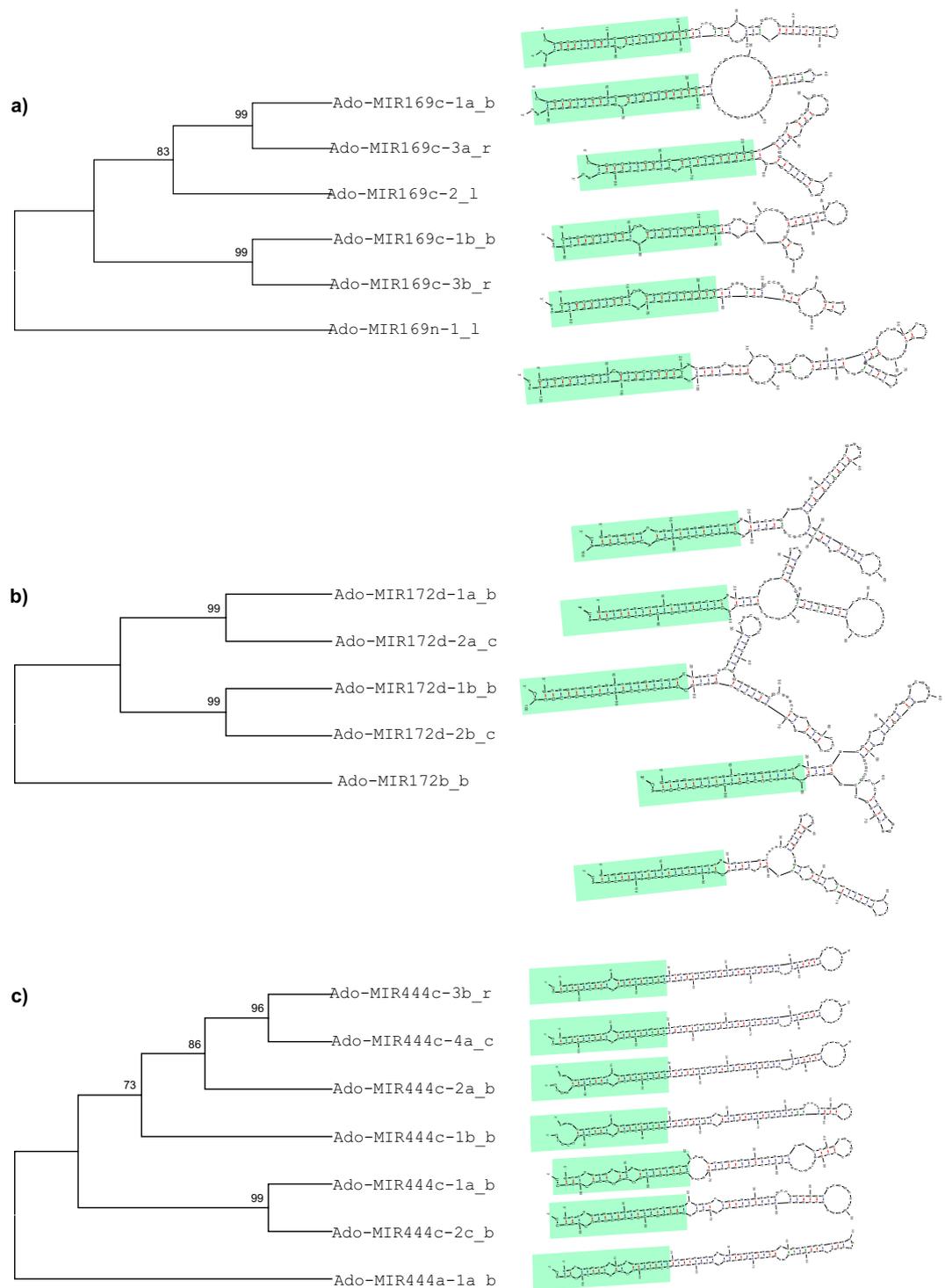


Figure 4. Stem-loop structures and phylogenetic relationships of *A. donax* pre-miRNA MIR169c (a), MIR172d (a) and MIR444c subfamilies (c). The single outgroups are at the base of the each cladogram. Uppercase letters identify mature microRNAs. Regions highlighted in light green are the stem structure of each pre-miRNA. Pre-miRNA structures were drawn with the C-mii software.

each family member and phylogenetic trees of corresponding family members were constructed using MIR169c, MIR172d and MIR444c multicopy loci as examples. As shown in Fig. 4, the stem structures were conserved among MIR169c (Fig. 4a) and MIR172d (Fig. 4b) paralogs, while divergent in MIR444c (Fig. 4c). On the contrary, the loop structures were conserved among MIR444c paralogs, while divergent in MIR169c and MIR172d subfamilies. These results were consistent with the phylogenetic relationships of MIR169c (Fig. 4a), MIR172d (Fig. 4b) and MIR444c (Fig. 4c) loci. In general, the more similar the miRNA structures were, the more likely they formed supported clades in the respective phylogenetic trees.

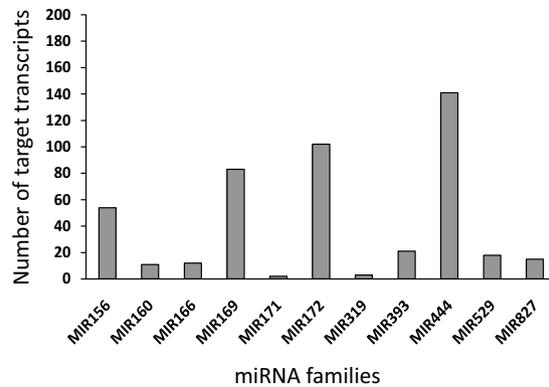


Figure 5. Number of predicted targets for potential *A. donax* miRNA families.

Target prediction of *A. donax* miRNAs and cleavage site validation. Despite the high sequence complementarity of plant microRNAs and their targets, different algorithms display marked differences in the total number of predicted targets from the same set of microRNAs/mRNA transcripts⁴⁷. For this reason we combined the predictions carried out with two popular programs, Targetfinder and psRNATarget, to attain a more reliable set of true positive targets⁴⁸. By considering only the common hits obtained with these two programs, and removed non-coding transcripts by sequences similarity search against known biological protein database, we predicted a total of 107 mature miRNAs out of 141 mature miRNAs to regulate with high reliability 462 non-redundant target transcripts in *A. donax* (Supplementary Fig. 3; Supplementary Table 5; Supplementary sequences). MIR156, MIR160, MIR166, MIR169, MIR171, MIR172, MIR319, MIR393, MIR444, MIR529 and MIR827 were observed to regulate more than one gene (Fig. 5). 102 out of 107 *A. donax* miRNAs were found to target more than one type of transcript, whereas Ado-MIR160b-2_r, Ado-MIR160b-3_r, Ado-MIR319a-1a_c, Ado-MIR393b-1a_b and Ado-MIR393b-2c_c only targeted single genes (Supplementary Fig. 4). miRNAs, therefore, tend to regulate multiple distinct genes, and their targets belong to several gene families which are involved in different biological process, cellular component and molecular function (Fig. 6). These results indicate that miRNAs in *A. donax* combinatorially control multiple processes by regulating different target genes during plant growth and development, as suggested for other species^{9,49}.

The type of post-transcriptional regulation was automatically analyzed for the predicted targets by the psRNATarget program. Out of 462 targets, 449 were predicted to undergo a cleavage type of inhibition. In particular, miRNA in the families MIR156, MIR160, MIR166, MIR169, MIR171, MIR172, MIR319, MIR393, MIR444, MIR529 and MIR827 showed cleavage type of inhibition for all predicted target transcripts. Besides the 449 target transcripts regulated by transcriptional cleavage, MIR444 was the only one predicted to regulate also a total of 13 transcripts by translational suppression. These results indicate that miRNA-mediated post-transcriptional gene silencing (PTGS) in *A. donax* took place mainly through cleavage of the target transcripts, while, in line with what previously observed in plants⁵⁰, translational regulation contributes marginally to overall microRNA regulatory activity.

To evaluate the method applied here for miRNA target prediction, four targets (one conserved and three not homologous to previously reported targets in other species) were used to determine the nucleotide position at which a cleavage event occurs using 5'-RACE by sequencing the amplified PCR products. Based on these analyses, the target cleavage sites reported in Fig. 7 were observed. The major cutting site of *Ado-MIR444a-1b_b:comp64193_c0_seq. 1* in buds was at the 11th position, while the target *comp88323_c0_seq. 2* of *Ado-MIR166a-2_r* in roots was cut with low efficiency and less precisely; The target *Locus_8503_Transcript_56/83_Confidence_0.566_Length_5268* of *Ado-MIR319a-2b_r* in roots was sliced in several positions including canonical sites, while the other *Ado-MIR444d-2c_b* new target in *Locus_10795_Transcript_53/70_Confidence_0.480_Length_2677* was cleaved preferentially in two positions in buds, but not in canonical sites.

Functional annotation of predicted targets. For functional annotation, gene ontology (GO) analysis was carried out for the predicted *A. donax* targets, which indicated their involvement in regulating diverse physiological processes (Supplementary Table 6). In line with the major role of microRNAs in the regulation of transcriptional cascades by targeting transcription factors⁵¹, the main molecular functions associated to the predicted *A. donax* target genes were binding and more specifically nucleic acid binding transcription factor activity, while only a minority of the functions referred to catalytic activity. For the cellular component category, the majority of the target genes were associated with cell part, organelle and macromolecular complex. Under the biological process category, the majority of targets were associated with biological regulation, developmental process and reproductive process. The functional annotation of the targets also was performed by sequence similarity searches against the *Arabidopsis thaliana* and *Setaria italica* proteins using the BLASTx algorithm. A total of 425 (91.99%) target genes functional hits were found in *Arabidopsis thaliana*, and a total of 455 (98.48%) in *Setaria italica*, both species with fully sequenced genomes.

Conservation of miRNA targets among *Arundo donax* and other plants. In order to investigate the conservation of the putative miRNA targets in *Arundo donax* compared to other known species namely *Oryza*

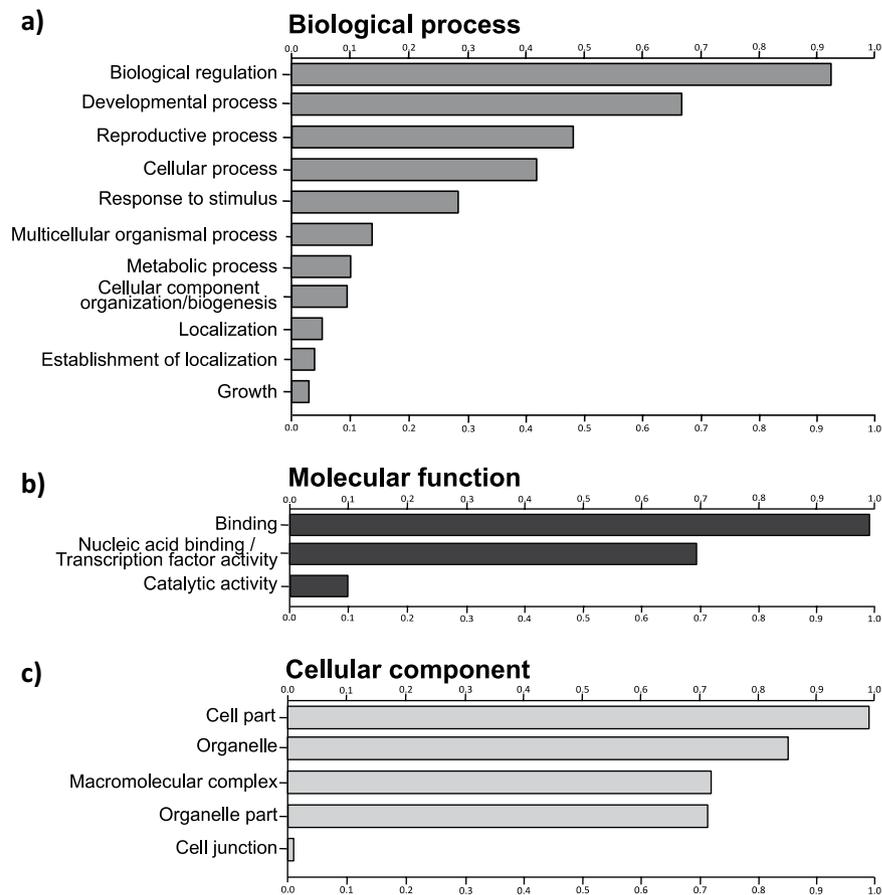


Figure 6. Functional annotation of predicted targets. **(a)** biological process, **(b)** molecular function and **(c)** cellular component.

sativa, *Zea mays*, *Arabidopsis thaliana* and *Vitis vinifera*, BLAST sequence similarity search approach was utilized to identify conserved miRNA targets. A total of 390 homologous targets for 10 conserved miRNA families were identified in *A. donax*, while 72 targets for eight *A. donax* conserved miRNA families were identified as novel targets (Table 2). Thus 84.42% of *A. donax* mRNA targets for the conserved miRNA families considered were conserved in *A. donax* and the other species. The remaining 15.58% of the conserved miRNAs targeted novel target transcripts.

Discussion

Identification of expressed microRNA loci from *A. donax*, an important biomass species, may have relevant applications to improve the quality/amount of its biomass, thanks to the establishment of promising protocols for regeneration and genetic transformation^{52,53}. In addition, it can significantly contribute to elucidate the evolutionary trajectories followed by such important regulators of gene expression in polyploid species²³. Transcriptome sequencing is one of the most efficient and cost-effective approaches available for gene discovery, and can provide massive and valuable information for the identification of low abundance or tissue-specific miRNA and their targets^{54–57}, phylogenetic inferences and characterization of polyploid speciation^{58,59}. In addition, the choice to leverage on the reference transcriptome of *A. donax* for microRNA identification has been dictated by the lack of a fully sequenced genome for this species. As such, we did not expect to be able to obtain a full representation of all microRNA loci existing in *A. donax*. We, thus, decided to carry out a stringent identification of the putative microRNA pri-miRNA by relying on a series of filtering steps. Especially the use of a threshold of -0.85 kcal/mol for MFEI provided a set of well resolved miRNA candidates with respect to tRNAs (average MFEI = -0.64), rRNAs (MFEI = -0.59) and mRNAs (MFEI = -0.65)^{12,46}. The length distributions of predicted pre- and mature-microRNAs (from 60 to 193 nt and from 20 to 22 nt, respectively) are in line with those observed in other species^{12,14,60,61}, thus supporting the reliability of the identification. Also the analysis of position-specific nucleotide preferences confirms similarity of *A. donax* mature miRNAs to those from other species. In particular, dominance of uracil at the first position of the 5' terminus may play an important role in miRNA biogenesis or RISC formation, while preference for cytosine at position 19 seems to be relevant for targeting RISC or Dicer-mediated cleavage to specific sites in pre-miRNAs^{60,62,63}. We note, however, that the set of identified microRNAs is not comprehensive and, therefore, there is a concrete possibility that bias in the nucleotide preference estimation may exist. Comparison of the number of microRNA loci present in the rice genome with the number of tentative loci identified in *A. donax* indicates that the major and evolutionarily most conserved families have been identified in

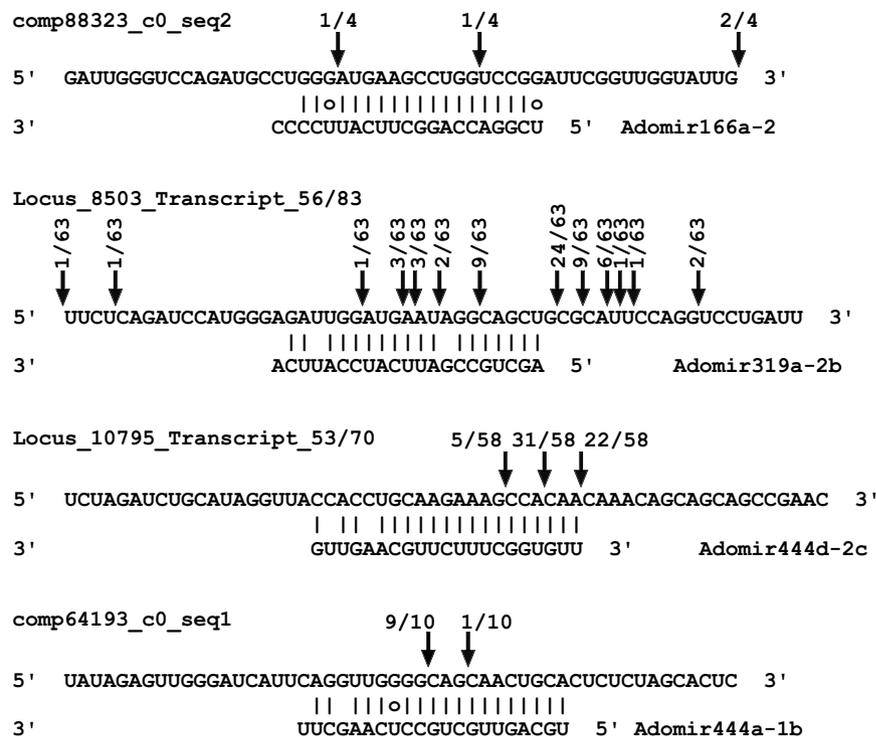


Figure 7. The cleavage site identification of four miRNA target genes with 5'-race in *A. donax*. The target gene is shown on top and corresponding miRNA on the bottom. The arrows indicate cleavage sites, and the numbers indicate the fraction of cloned PCR products which were used to determine the cleavage sites.

miRNA families	<i>A. donax</i>	<i>O. sativa</i>	<i>Z. mays</i>	<i>A. thaliana</i>	<i>V. vinifera</i>	Novel targets
MIR156	50	48	45	40	47	4
MIR160	11	10	9	9	10	0
MIR166	11	11	11	11	11	1
MIR169	77	75	77	68	58	6
MIR171	2	2	2	2	2	0
MIR172	74	74	74	51	60	28
MIR319	0	0	0	0	0	3
MIR393	19	19	19	19	19	2
MIR444	120	118	110	—	—	21
MIR529	11	10	1	—	—	7
MIR827	15	15	15	0	—	0
Total	390	382	363	200	207	72

Table 2. Numbers of *A. donax* conserved miRNA targets and novel targets in conserved miRNA families across representative monocots and dicots plant genomes. Note: “—” represent the miRNA targets absence; “0” indicates lack of the corresponding miRNA in the species.

our screening⁶³. The relatively higher numbers of loci identified for families MIR444 and MIR1430 further suggest that lineage-specific expansion of these microRNAs possibly took place in *A. donax* after its divergence from the rice lineage, occurred approximately 80 million years ago⁶⁴. While the involvement of MIR444 in mediating perception of viral infections is well established (see below), the function of MIR1430 remains to be fully elucidated. Previous studies indicate that in rice MIR1430 post-transcriptionally regulates a gene encoding a member of the nuclear factor Y-A (NF-YA) gene family^{65,66} (LOC_Os12g42400). The transcript of LOC_Os12g42400 is expressed specifically in callus, flower and panicles⁶⁷ and it is co-regulated by MIR169 family members⁶⁵. In Arabidopsis MIR169 takes part in stress-induced early flowering⁶⁸. MIR1430 family expansion may predate *A. donax* loss of fertility associated to polyploidization⁶⁹, when transition to the reproductive phase likely still constituted a relevant stress-escape strategy for this species. Another possibility is that the likely lineage-specific expansion of MIR1430 gene family may be related to suppression of fungal resistance, as differential repression of NF-YA genes by MIR169 has been found to negatively regulate rice immunity against the blast fungus *M. oryzae*⁷⁰.

Validation of four target genes confirmed in all cases slicing within the target region, albeit with different efficiency and specificity. Previous studies demonstrated that the cleavage sites of miRNA targets do not only occur at the canonical position between 10th and 11th nucleotides (nt) from the 5' end of the miRNA and miRNA:target base pairing regions, but also happen outside of base pairing regions^{71,72}. The conserved targets comp88323_c0_seq.2 of MIR166a, for instance, had diverse cleavage sites compared to the canonical position in its *Oryza sativa* orthologs^{72,73}. This is likely due to the highly dynamic nature of microRNA slicing sites, as even the same microRNA can cleave its target in different places in different physiological conditions⁷⁴. In addition, cleavage site position has been documented to vary among species for the same miRNA/target combination⁷⁴. Worth of note, the other three targets we validated are specific for *A. donax*. Other microRNA/target pairs have been demonstrated to be species-specific⁷², indicating that they contribute to interspecific differences in gene regulation. Our results, thus suggest that in addition to the functions already demonstrated in rice (tillering control⁷⁵; antiviral defence⁷⁶), based on the function of Arabidopsis target orthologs AT3G49170 and AT1G13640, *A. donax* MIR444a and MIR444d could play a role also in additional cellular mechanisms, like, respectively, chloroplast RNA editing and acyl lipid metabolism^{77,78}. Analogously, MIR319 is possibly involved in regulation of post-Golgi trafficking by targeting the orthologue of Arabidopsis AT3G60860, an ARF guanine-nucleotide exchange factor⁷⁹. While as expected the large majority of miRNA/target mRNA pairs are conserved in other angiosperm species (Table 2), our results indicate that up to 15% of the targets of the highly conserved microRNA families may be novel in *A. donax* and/or other species.

In line with the known tendency of microRNAs to preferentially target transcription factors⁸⁰, we found that the major molecular function classes targeted in *A. donax* were “binding” and “nucleic acid binding transcription factor activity”. Also the most common biological process GO classes indicate that the sets of microRNAs identified in *A. donax* play an important role in the regulation of development, reproduction, metabolism as well as stress response. In particular, network analysis of *A. donax* miRNAs and their target genes highlighted the large number of targets regulated by MIR444, MIR172, MIR169 and MIR156. In *Oryza sativa*, RNA-dependent RNA polymerase 1 (RDR1) is a central component in the antiviral RNA-silencing pathway, and MIR444 plays an important role in transducing the antiviral signal from virus infection to RDR1 expression⁷⁶. The capacity of *A. donax* to asymptotically stand viral infections may, thus, in part depend on the possible expansion of the MIR444 family in this species⁸¹. Previous studies showed that MIR156 regulates developmental timing by repressing the expression of functionally distinct SPL transcription factors, while MIR172 regulates flowering time and flower formation by regulating the expression of AP2-like transcription factors^{82,83}. MIR172 could, therefore, be a useful tool to modulate flowering time in *A. donax*, analogously to what observed in *Sorghum*⁸⁴. The MIR169/NF-YA (Nuclear factor Y, subunit A) is a well established regulatory module functioning in developmental processes and responding to environmental stresses^{85,86}. Also MIR166 may possibly respond to environmental stress and it controls root architecture⁸⁷, a trait that may be relevant for improving *A. donax* tolerance to drought³⁴.

The growing interest towards improvement of biofuel/bioenergy crops has stimulated in recent years the search for novel approaches to improve their productivity. As specific miRNAs regulate several bioenergy traits, genetic transformation of bioenergy crops like switchgrass and poplar with selected microRNAs has been already demonstrated a viable option to improve plant biomass, decreasing the lignin content, modulating stress responses and flowering time^{88,89}. This is, however, still just a small fraction of the 30 microRNA families of potential interest for bioenergy crop improvement⁹⁰. Dissecting the genetic architecture of miRNA loci in the crop of interest is the first fundamental step for any subsequent attempt to improve biofuel feedstock species⁹⁰. The putative microRNA loci identified in the present study from the transcriptome of *A. donax* provide novel opportunities for the genetic improvement of biomass yield and quality in this emerging biomass species. They also shed new light into the complex dynamics of microRNA evolution in this highly polyploid species, providing evidence for lineage-specific amplification of microRNA families and targeting of novel cellular mRNAs involved in important aspects of plant fitness and productivity, like RNA editing, acyl lipid metabolism and post-Golgi trafficking. In plants several non-conserved, species-specific microRNAs families have been found⁹¹. As mining of species-specific miRNAs has still not been carried out in *A. donax*, it seems likely that the number of microRNA targets yet to be identified could be significantly larger than that estimated in our study. In-depth identification of such novel targets will thus be required to attain a more complete understanding of microRNA regulation in this important bioenergy crop.

Data Availability

All data analysed during this study are included in this article (and its Supplementary Information files).

References

1. Bartel, D. P. MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell* **116**, 281–297 (2004).
2. Lee, Y. *et al.* MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.* **23**, 4051–4060 (2004).
3. Kurihara, Y. & Watanabe, Y. From The Cover: Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions. *Proc. Natl. Acad. Sci.* **101**, 12753–12758 (2004).
4. Davis, B. N. & Hata, A. Regulation of MicroRNA Biogenesis: A miRiad of mechanisms. *Cell Commun. Signal.* **7**, 1–22 (2009).
5. Baumberger, N. & Baulcombe, D. C. Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. *Proc. Natl. Acad. Sci.* **102**, 11928–11933 (2005).
6. German, M. A. *et al.* Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. *Nat. Biotechnol.* **26**, 941–946 (2008).
7. Kidner, C. A. & Martienssen, R. A. The developmental role of microRNA in plants. *Curr. Opin. Plant Biol.* **8**, 38–44 (2005).
8. Zhang, B., Pan, X., Cobb, G. P. & Anderson, T. A. Plant microRNA: A small regulatory molecule with big impact. *Dev. Biol.* **289**, 3–16 (2006).

9. Dehury, B. *et al.* In silico identification and characterization of conserved miRNAs and their target genes in sweet potato (*Ipomoea batatas* L.) Expressed Sequence Tags (ESTs). *Plant Signal. Behav.* **8**, e26543 (2013).
10. Singh, N., Srivastava, S. & Sharma, A. Identification and analysis of miRNAs and their targets in ginger using bioinformatics approach. *Gene* **575**, 570–576 (2016).
11. Lindow, M. & Krogh, A. Computational evidence for hundreds of non-conserved plant microRNAs. *BMC Genomics* **6**, 1–9 (2005).
12. Patanun, O., Lertpanyasampatha, M., Sojikul, P., Viboonjun, U. & Narangajavana, J. Computational identification of MicroRNAs and their targets in cassava (*Manihot esculenta* Crantz.). *Mol. Biotechnol.* **53**, 257–269 (2013).
13. Dong, Q. H. *et al.* Computational identification of microRNAs in strawberry expressed sequence tags and validation of their precise sequences by miR-RACE. *J. Hered.* **103**, 268–277 (2012).
14. Wang, X.-J., Reyes, J. L., Chua, N.-H. & Gaasterland, T. Prediction and identification of Arabidopsis thaliana microRNAs and their mRNA targets. *Genome Biol.* **5**, R65 (2004).
15. Archak, S. & Nagaraju, J. Computational Prediction of Rice (*Oryza sativa*) miRNA Targets. *Genomics, Proteomics Bioinforma.* **5**, 196–206 (2007).
16. Griffiths-Jones, S., Saini, H. K., Van Dongen, S. & Enright, A. J. *miRBase: Tools for microRNA genomics.* *Nucleic Acids Res.* **36**, 154–158 (2008).
17. Yin, Z., Li, C., Han, X. & Shen, F. Identification of conserved microRNAs and their target genes in tomato (*Lycopersicon esculentum*). *Gene* **414**, 60–66 (2008).
18. Bonnet, E., Wuyts, J., Rouzé, P. & Van de Peer, Y. Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences. *Bioinformatics* **20**, 2911–2917 (2004).
19. Rhoades, M. W. *et al.* Prediction of plant microRNA targets. *Cell* **110**, 513–520 (2002).
20. Schwab, R. *et al.* Specific effects of microRNAs on the plant transcriptome. *Dev. Cell* **8**, 517–527 (2005).
21. Devi, K. J., Chakraborty, S., Deb, B. & Rajwanshi, R. Computational identification and functional annotation of microRNAs and their targets from expressed sequence tags (ESTs) and genome survey sequences (GSSs) of coffee (*Coffea arabica* L.). *Plant Gene* **6**, 30–42 (2016).
22. Rossa, B., Tuffers, A. V., Naidoo, G. & von Willert, D. J. *Arundo donax* L. (Poaceae) - a C-3 species with unusually high photosynthetic capacity. *Bot. ACTA* **111**, 216–221 (1998).
23. Hardion, L., Verlaque, R., Saltonstall, K., Leriche, A. & Vila, B. Origin of the invasive *Arundo donax* (Poaceae): a trans-Asian expedition in herbaria. *Ann. Bot.* mcl43, <https://doi.org/10.1093/aob/mcu143> (2014).
24. Hardion, L., Verlaque, R., Baumel, A., Juin, M. & Vila, B. Revised systematics of Mediterranean *Arundo* (Poaceae) based on AFLP fingerprints and morphology. *Taxon* **61**, 1217–1226 (2012).
25. Bucci, A., Cassani, E., Landoni, M., Cantaluppi, E. & Pilu, R. Analysis of chromosome number and speculations on the origin of *Arundo donax* L. (Giant Reed). *Cytol. Genet.* **47**, 237–241 (2013).
26. Hoshovsky, M. *Arundo donax.* *Element Stewardship Abstract.* **1988** (1987).
27. Angelini, L. G., Ceccarini, L., Nassi o Di Nasso, N. & Bonari, E. Comparison of *Arundo donax* L. and *Miscanthus x giganteus* in a long-term field experiment in Central Italy: Analysis of productive characteristics and energy balance. *Biomass and Bioenergy* **33**, 635–643 (2009).
28. Calheiros, C. S. C. *et al.* Use of constructed wetland systems with *Arundo* and *Sarcocornia* for polishing high salinity tannery wastewater. *J. Environ. Manage.* **95**, 66–71 (2012).
29. Papazoglou, E. G., Karantounias, G. A., Vemmos, S. N. & Bouranis, D. L. Photosynthesis and growth responses of giant reed (*Arundo donax* L.) to the heavy metals Cd and Ni. *Environ. Int.* **31**, 243–249 (2005).
30. Raspolli Galletti, A. M. *et al.* From giant reed to levulinic acid and gamma-valerolactone: A high yield catalytic route to valeric biofuels. *Appl. Energy* **102**, 157–162 (2013).
31. Scordia, D., Cosentino, S. L., Lee, J.-W. & Jeffries, T. W. Dilute oxalic acid pretreatment for biorefining giant reed (*Arundo donax* L.). *Biomass Bioenergy* **35**, 3018–3024 (2011).
32. Sablok, G. *et al.* Fuelling genetic and metabolic exploration of C3 bioenergy crops through the first reference transcriptome of *Arundo donax* L. *Plant Biotechnol. J.* **12**, 554–567 (2014).
33. Barrero, R. A. *et al.* Shoot transcriptome of the giant reed, *Arundo donax*. *Data Br.* **3**, 1–6 (2015).
34. Fu, Y. *et al.* Dissection of early transcriptional responses to water stress in *Arundo donax* L. by unigene-based RNA-seq. *Biotechnol. Biofuels* **9**, 54 (2016).
35. Kozomara, A. & Griffiths-Jones, S. MiRBase: Annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* **42**, 68–73 (2014).
36. Numnark, S., Mhuantong, W., Ingriswang, S. & Wichadakul, D. C-mii: a tool for plant miRNA and target identification. *BMC Genomics* **13**(Suppl 7), S16 (2012).
37. Prakash, P., Ghosliya, D. & Gupta, V. Identification of conserved and novel microRNAs in *Catharanthus roseus* by deep sequencing and computational prediction of their potential targets. *Gene* **554**, 181–195 (2015).
38. Xu, J. H., Li, F. & Sun, Q. F. Identification of MicroRNA Precursors with Support Vector Machine and String Kernel. *Genomics, Proteomics Bioinforma.* **6**, 121–128 (2008).
39. Hall, T. BioEdit: An important software for molecular biology. *GERF Bull. Biosci.* **2**, 60–61 (2011).
40. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).
41. Dai, X. & Zhao, P. X. PsRNATarget: A plant small RNA target analysis server. *Nucleic Acids Res.* **39**, 155–159 (2011).
42. Xiaochen, B. & Wang, S. TargetFinder: A software for antisense oligonucleotide target site selection based on MAST and secondary structures of target mRNA. *Bioinformatics* **21**, 1401–1402 (2005).
43. Chen, T. W. *et al.* FunctionAnnotator, a versatile and efficient web tool for non-model organism annotation. *Sci. Rep.* **7**, 1–9 (2017).
44. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2498–2504** <https://doi.org/10.1101/gr.1239303.metabolite> (2003).
45. Adai, A. *et al.* Computational prediction of miRNAs in *Arabidopsis thaliana*. *Genome Res.* **15**, 78–91 (2005).
46. Zhang, B. H., Pan, X. P., Cox, S. B., Cobb, G. P. & Anderson, T. A. Evidence that miRNAs are different from other RNAs. *Cell. Mol. Life Sci.* **63**, 246–254 (2006).
47. Yue, D., Liu, H. & Huang, Y. Survey of Computational Algorithms for MicroRNA Target Prediction. *Curr. Genomics* **10**, 478–92 (2009).
48. Srivastava, P. K., Moturu, T. R., Pandey, P., Baldwin, I. T. & Pandey, S. P. A comparison of performance of plant miRNA target prediction tools and the characterization of features for genome-wide target prediction. *BMC Genomics* **15**, 1–15 (2014).
49. Yuan, X. *et al.* Clustered microRNAs' coordination in regulating protein-protein interaction network. *BMC Syst. Biol.* **3**, 1–10 (2009).
50. Sunkar, R., Girke, T., Jain, P. K. & Zhu, J.-K. Cloning and Characterization of MicroRNAs from Rice. *Plant Cell* **17**, 1397–1411 (2005).
51. Nazarov, P. V. *et al.* Interplay of microRNAs, transcription factors and target genes: Linking dynamic expression changes to function. *Nucleic Acids Res.* **41**, 2817–2831 (2013).
52. Takahashi, W., Takamizo, T., Kobayashi, M. & Ebina, M. Plant regeneration from calli in giant reed (*Arundo donax* L.). *Grassl. Sci.* **56**, 224–229 (2010).

53. Dhir, S., Knowles, K. & Pagan, C. L. Optimization and transformation of *Arundo donax* L. using particle bombardment. *African J. Biotechnol.* **9**, 6460–6469 (2010).
54. Xu, Y. *et al.* Transcriptome-wide identification of miRNAs and their targets from *Typha angustifolia* by RNA-Seq and their response to cadmium stress. *PLoS One* **10**, 1–22 (2015).
55. Ling, L. Z. *et al.* Transcriptome-wide identification and prediction of mirnas and their targets in paris polyphylla var. *Yunnanensis* by high-throughput sequencing analysis. *Int. J. Mol. Sci.* **18**, 1–12 (2017).
56. Chen, M., Bao, H., Wu, Q. & Wang, Y. Transcriptome-wide identification of miRNA targets under nitrogen deficiency in populus tomentosa using degradome sequencing. *Int. J. Mol. Sci.* **16**, 13937–13958 (2015).
57. Fahlgren, N. *et al.* High-throughput sequencing of Arabidopsis microRNAs: Evidence for frequent birth and death of MIRNA genes. *PLoS One* **2** (2007).
58. Wang, K., Hong, W., Jiao, H. & Zhao, H. Transcriptome sequencing and phylogenetic analysis of four species of luminescent beetles. *Sci. Rep.* **7**, 1–11 (2017).
59. Buggs, R. J. A. *et al.* Next-generation sequencing and genome evolution in allopolyploids. *Am. J. Bot.* **99**, 372–382 (2012).
60. Zhang, B., Pan, X. & Stellwag, E. J. Identification of soybean microRNAs and their targets. *Planta* **229**, 161–182 (2008).
61. Xuan, P. *et al.* PlantMiRNAPred: Efficient classification of real and pseudo plant pre-miRNAs. *Bioinformatics* **27**, 1368–1376 (2011).
62. Prakash, P., Rajakani, R. & Gupta, V. Transcriptome-wide identification of Rauwolfia serpentina microRNAs and prediction of their potential targets. **61**, 62–74 (2016).
63. Zhang, B., Pan, X., Cannon, C. H., Cobb, G. P. & Anderson, T. A. Conservation and divergence of plant microRNA genes. *Plant J.* **46**, 243–259 (2006).
64. Prasad, V. *et al.* Late Cretaceous origin of the rice tribe provides evidence for early diversification in Poaceae. *Nat. Commun.* **2**, 1–9 (2011).
65. Sun, W. *et al.* Genome-wide identification of microRNAs and their targets in wild type and phyB mutant provides a key link between microRNAs and the phyB-mediated light signaling pathway in rice. *Front. Plant Sci.* **6**, 1–15 (2015).
66. Thirumurugan, T., Ito, Y., Kubo, T., Serizawa, A. & Kurata, N. Identification, characterization and interaction of HAP family genes in rice. *Mol. Genet. Genomics* **279**, 279–289 (2008).
67. Liu, Q. *et al.* Integrating Small RNA Sequencing with QTL Mapping for Identification of miRNAs and Their Target Genes Associated with Heat Tolerance at the Flowering Stage in Rice. *Front. Plant Sci.* **8**, 1–15 (2017).
68. Xu, M. Y. *et al.* Stress-induced early flowering is mediated by miR169 in Arabidopsis thaliana. *J. Exp. Bot.* **65**, 89–101 (2014).
69. Hardion, L., Verlaque, R., Rosato, M., Rosselló, J. & Vila, B. Impact of polyploidy on fertility variation of Mediterranean *Arundo L.* (Poaceae). *C. R. Biol.* **338**, 298–306 (2015).
70. Li, Y. *et al.* Osa-miR169 Negatively Regulates Rice Immunity against the Blast Fungus *Magnaporthe oryzae*. *Front. Plant Sci.* **8**, 1–13 (2017).
71. Moxon, S. *et al.* Deep sequencing of tomato short RNAs identifies microRNAs targeting genes involved in fruit ripening. *Genome Res.* **18**, 1602–1609 (2008).
72. Li, Y. F. *et al.* Transcriptome-wide identification of microRNA targets in rice. *Plant J.* **62**, 742–759 (2010).
73. Yi, R. *et al.* Identification and Expression Analysis of microRNAs at the Grain Filling Stage in Rice (*Oryza sativa* L.) via Deep Sequencing. *PLoS One* **8** (2013).
74. Zeng, C. *et al.* Conservation and divergence of microRNAs and their functions in Euphorbiaceous plants. *Nucleic Acids Res.* **38**, 981–995 (2009).
75. Guo, S. *et al.* The interaction between OsMADS57 and OsTB1 modulates rice tillering via DWARF14. *Nat. Commun.* **4**, 1512–1566 (2013).
76. Wang, H. *et al.* A Signaling Cascade from miR444 to RDR1 in Rice Antiviral RNA Silencing Pathway. *Plant Physiol.* **170**, 2365–2377 (2016).
77. Mueller-Roeber, B. & Pical, C. Inositol Phospholipid Metabolism in Arabidopsis. Characterized and Putative Isoforms of Inositol Phospholipid Kinase and Phosphoinositide-Specific Phospholipase C. *Plant Physiol.* **130**, 22–46 (2002).
78. Sun, Y. K., Gutmann, B., Yap, A., Kindgren, P. & Small, I. Editing of Chloroplast rps14 by PPR Editing Factor EMB2261 Is Essential for Arabidopsis. *Development*. **9**, 1–11 (2018).
79. Kitakura, S. *et al.* BEN3/BIG2 ARF GEF is Involved in Brefeldin A-Sensitive Trafficking at the trans-Golgi Network/Early Endosome in Arabidopsis thaliana. *Plant Cell Physiol.* **58**, 1801–1811 (2017).
80. Cui, Q., Yu, Z., Pan, Y., Purisima, E. O. & Wang, E. MicroRNAs preferentially target the genes with high transcriptional regulation complexity. *Biochem. Biophys. Res. Commun.* **352**, 733–738 (2007).
81. Ingwell, L. L., Zemetra, R., Mallory-Smith, C. & Bosque-Pérez, N. A. *Arundo donax* infection with Barley yellow dwarf virus has implications for biofuel production and non-managed habitats. *Biomass and Bioenergy* **66**, 426–433 (2014).
82. Wu, G. *et al.* The sequential actions of miR156 and miR172 regulates developmental timing in Arabidopsis. *Cell* **138**, 750–759 (2009).
83. Zhu, Q. H. & Helliwell, C. A. Regulation of flowering time and floral patterning by miR172. *J. Exp. Bot.* **62**, 487–495 (2011).
84. Calviño, M., Bruggmann, R. & Messing, J. Characterization of the small RNA component of the transcriptome from grain and sweet sorghum stems. *BMC Genomics* **12**, 356 (2011).
85. Sorin, C. *et al.* A miR169 isoform regulates specific NF-YA targets and root architecture in Arabidopsis. *New Phytol.* **202**, 1197–1211 (2014).
86. Li, Y., Fu, Y., Ji, L., Wu, C. & Zheng, C. Characterization and expression analysis of the Arabidopsis mir169 family. *Plant Sci.* **178**, 271–280 (2010).
87. Boualem, A. *et al.* MicroRNA166 controls root and nodule development in *Medicago truncatula*. *Plant J.* **54**, 876–887 (2008).
88. Fu, C. *et al.* Overexpression of miR156 in switchgrass (*Panicum virgatum* L.) results in various morphological alterations and leads to improved biomass production. *Plant Biotechnol. J.* **10**, 443–52 (2012).
89. Rubinelli, P. M., Chuck, G., Li, X. & Meilan, R. Constitutive expression of the Corngrass1 microRNA in poplar affects plant architecture and stem lignin content and composition. *Biomass and Bioenergy* **54**, 312–321 (2013).
90. Trumbo, J. L., Zhang, B. & Stewart, C. N. Manipulating microRNAs for improved biomass and biofuels from plant feedstocks. *Plant Biotechnol. J.* **13**, 337–354 (2015).
91. Qin, Z., Li, C., Mao, L. & Wu, L. Novel insights from non-conserved microRNAs in plants. *Front. Plant Sci.* **5**, 1–10 (2014).

Acknowledgements

This work was supported by the Autonomous Province of Trento (Italy) and by the China Scholarship Council (JW: 201406300062).

Author Contributions

W.J. carried out the bioinformatic analyses and drafted the MS. G.S. helped with the bioinformatics analyses and participated to MS correction. G.B. participated to the supervision of the study and participated to MS correction. M.L. conceived the study, supervised it, reviewed and finalized the MS. C.V. conceived the study, supervised it, reviewed and finalized the MS.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-34982-8>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018